From MEDIEVAL ROBOTS to NEURAL NETWORKS

ARTIFICIAL INTELLIGENCE

AN ILLUSTRATED HISTORY

CLIFFORD A. PICKOVER

VISIT…

LANZAROTE Caliente.COM

BOOKS BY CLIFFORD A. PICKOVER

The Alien IQ Test
Archimedes to Hawking
A Beginner's Guide to Immortality
Black Holes: A Traveler's Guide
The Book of Black
Brain Strain: A Mental Muscle Workout That's Fun!
Calculus and Pizza
Chaos and Fractals (ed.)
Chaos in Wonderland
Computers and the Imagination
Computers, Pattern, Chaos, and Beauty
Cryptorunes: Codes and Secret Writing
Death and the Afterlife
Dreaming the Future
Egg Drop Soup
Fractal 3D Magic
Future Health (ed.)
Fractal Horizons: The Future Use of Fractals (ed.)
Frontiers of Scientific Visualization (ed.)
The Girl Who Gave Birth to Rabbits
The Heaven Virus
Jews in Hyperspace
Keys to Infinity
Liquid Earth
The Lobotomy Club

The Loom of God

The Math Book

The Mathematics Devotional

The Mathematics of Oz

Mazes for the Mind: Computers and the Unexpected

The Medical Book

The Möbius Strip

The Paradox of God and the Science of Omniscience

A Passion for Mathematics

The Pattern Book: Fractals, Art, and Nature (ed.)

The Physics Book

The Physics Devotional

The Science Book

The Science of Aliens

Sex, Drugs, Einstein, and Elves

Spider Legs (with Piers Anthony)

Spiral Symmetry (ed., with Istvan Hargittai)

The Stars of Heaven

Strange Brains and Genius

Surfing through Hyperspace

Sushi Never Sleeps

Time: A Traveler's Guide

Visions of the Future (ed.)

Visualizing Biological Information (ed.)

Wonders of Numbers

The Zen of Magic Squares, Circles, and Stars

ARTIFICIAL INTELLIGENCE

AN ILLUSTRATED HISTORY

From MEDIEVAL ROBOTS to NEURAL NETWORKS

CLIFFORD A. PICKOVER

STERLING
New York

STERLING
New York

An Imprint of Sterling Publishing Co., Inc.
1166 Avenue of the Americas
New York, NY 10036

ISBN 978-1-4549-3360-1

For information about custom editions, special sales, and premium and corporate purchases, please contact Sterling Special Sales at 800-805-5489 or specialsales@sterlingpublishing.com.

sterlingpublishing.com

Cover and endpaper design by Elizabeth Mihaltse Lindy
Front cover illustration by Matthieu Bourel (WFH)
For image credits, see page 206

“An attempt will be made to find how to make machines use language, form abstractions and concepts, solve the kinds of problems now reserved for humans, and improve themselves. . . . For the present purpose, the artificial intelligence problem is taken to be that of making a machine behave in ways that would be called intelligent if a human were so behaving.”

—John McCarthy, Marvin Minsky, Nathaniel Rochester, and Claude Shannon, “Proposal for the Dartmouth Summer Research Project on Artificial Intelligence,” 1955

“Artificial Intelligence can drive cars, trade stocks and shares, learn to carry out complex skills simply by watching YouTube videos, translate across dozens of different languages, recognize human faces with more accuracy than we can, and create original hypotheses to help discover new drugs for curing disease. That’s just the beginning.”

—Luke Dormehl, *Thinking Machines*, 2017

“Not until a machine can write a sonnet or compose a concerto because of thoughts and emotions felt, and not by the chance fall of symbols, could we agree that machine equals brain—that is, not only write it but know that it had written it.”

—Professor Geoffrey Jefferson, “The Mind of Mechanical Man,” 1949

“Whether we are based on carbon or on silicon makes no fundamental difference; we should each be treated with appropriate respect.”

—Arthur C. Clarke, *2010: Odyssey Two*, 1984

“Emerging as it does from many fields—philosophy, mathematics, psychology and even neurology—AI raises basic questions about human intelligence, memory, the

mind/body problem, the origins of language, symbolic reasoning, information processing, and so forth. AI researchers are—like alchemists of old who sought to create gold from base metal—seeking to create thinking machines from infinitesimal small bits of silicon oxyde."

—Daniel *Crevier*, *AI: The Tumultuous History of the Search for Artificial Intelligence,* 1993

CONTENTS

1774: Jaquet-Droz Automata

1818: *Frankenstein*

1821: Computational Creativity

1822: Babbage's Mechanical Computer

1844: "The Artist of the Beautiful"

1854: Boolean Algebra

1863: "Darwin among the Machines"

1868: *The Steam Man of the Prairies*

1883: Tower of Hanoi

1893: *Electric Bob's Big Black Ostrich*

1898: Tesla's "Borrowed Mind"

1907: Tik-Tok

1907: Searches for the Soul

1913: Optical Character Recognition (OCR)

1920: *Rossum's Universal Robots*

1927: *Metropolis*

1939: Elektro the Moto-Man

1939: Speech Synthesis

1942: Asimov's Three Laws of Robotics

1942: Lethal Military Robots

1943: Artificial Neural Networks

1946: ENIAC

1949: *Giant Brains, or Machines That Think*

1950: Turing Test

INTRODUCTION

"The period of time occupied by organic intelligence is just a thin sliver between early life and the long era of the machines."

—Martin Rees, *The Conversation*, April 2017 interview

AI and Beyond

"A lot of cutting edge AI has filtered into general applications, often without being called AI, because once something becomes useful enough and common enough, it's not labeled AI anymore."

—Nick Bostrom, "AI Set to Exceed Human Brain Power," *CNN.com*, 2006

Throughout history, the mysteries of the mind, the nature of thought, and the possibility of artificial beings have captivated artists, scientists, philosophers, and even theologians. Symbols and stories involving *automata*—moving mechanical devices made in imitation of living beings—permeate myth, art, music, and literature. Our fascination with artificial intelligence (AI)—apparently intelligent behavior by machines—is also reflected in the spooky or transcendent themes of blockbuster films or video games involving emotional robots and advanced intelligences that we can barely comprehend.

In this book, we'll embark on a vast chronological journey from ancient games to advanced modern computing approaches involving artificial neural networks that learn and improve their performance, often with little or no task-specific programming and rules. Along the way, we'll encounter odd and perplexing marvels like the mysterious

copper knights of Arthurian legend. We'll also encounter French inventor Jacques de Vaucanson's Canard Digérateur, a hyper-realistic duck automaton that inspired American author Thomas Pynchon's historical novel *Mason & Dixon* more than 250 years later, and the thirteenth-century Catalan philosopher Ramon Llull, who was among the first to explore a systematic approach to artificially generating ideas using a mechanical device. Skip forward to 1893, and we encounter the quirky and entertaining *Electric Bob's Big Black Ostrich*, a story that—along with the series *The Steam Man of the Prairies*—is notable for reflecting the growing fervor for all things mechanical during the Victorian-era steampunk movement.

In more recent times, we encounter Arthur Samuel of IBM, who implemented one of the earliest computer programs for playing checkers in 1952, followed in 1955 by a program that *learned* to play the game without outside interference. Today, the term *artificial intelligence* often refers to systems designed to learn, solve problems, and interact with humans using natural-language processing. Intelligent personal assistants like Amazon's Alexa, Apple's Siri, and Microsoft's Cortana all reflect some aspect of AI.

In this book, we'll also address fascinating issues concerning the ethical use of AI and even the challenges of placing advanced AI entities, should they become dangerously super-intelligent, into a "leakproof" box in order to isolate them from the outside world. Of course, the boundaries and scope of AI morph through time, and some experts suggest inclusive definitions that admit a range of technologies that have helped human beings perform cognitive tasks. Thus, to give a richer understanding of AI history, I have also included several devices or machines that have provided answers to problems that typically require human thinking and human computation, including the abacus, the Antikythera Mechanism (125 BCE), ENIAC (1946), and more. After all, without these early technologies, we would not have the advanced chess-playing and vehicle-driving systems that exist in our modern world.

As you read this book, remember that even if we consider some of the historical ideas or predictions concerning artificial beings to be far-fetched, old ideas may become suddenly viable when

implemented on faster, more advanced computer hardware. Our technical predictions—and even our myths—are, at a minimum, fascinating models of human understanding and creativity—and of how we reach across cultures and time to understand one another and learn about what we hold sacred or beneficial for society. However, even as we celebrate human imagination and ingenuity, it is vital to discuss unintended consequences, including the possible dangers of AI. As theoretical physicist Stephen Hawking told the BBC in a 2014 interview, “The development of full artificial intelligence could spell the end of the human race. . . . It would take off on its own, and redesign itself at an ever-increasing rate.” In other words, there’s a chance that AI entities will become so smart and capable that they will be able to continually improve themselves, creating a kind of superintelligence that could pose a great risk to humanity. This form of runaway technology growth, sometimes referred to as the *technological singularity*, could result in unimaginable changes to civilization, society, and human life.

So while the potential benefits of AI are numerous—self-driving cars, efficient business processes, and even companionship in countless arenas—humanity will need to be particularly cautious when developing autonomous weapons systems and over-relying on AI technologies with sometimes inscrutable mechanisms. For example, studies show the ease with which some AI (neural net) imaging systems can be “tricked” into misidentifying animals as rifles, or a picture of a plane as a dog, by altering images in ways that humans cannot perceive. If a terrorist can make a mall or a hospital look like a military target to a drone, the consequences could be dire. On the other hand, perhaps armed machines with appropriate sensors and rules of ethics could also reduce civilian casualties. Informed policy-making is needed to ensure that the potential dangers of AI entities do not overshadow their amazing benefits.

As we increasingly put our trust in AIs with many complex deep learning neural networks, one interesting area of research is developing AI systems that can *explain* to humans how they arrived at certain decisions. However, forcing AIs to explain themselves

could potentially cripple them, at least in certain applications. Many of these machines can create far more intricate models of reality than humans could possibly understand. AI expert David Gunning has even suggested that the highest-performing system will be the least explainable.

Book Organization and Purpose

I have had a longtime fascination with computing and topics at the borderlands of science, and my goal in writing this book is to provide a wide audience with a brief guide to both curious and important practical ideas in the history of *artificial intelligence*, a term that wasn't coined until 1955 by computer scientist John McCarthy. Each book entry is only a few paragraphs in length, allowing readers to jump in at any point in the book without having to sort through a lot of verbiage. Of course, this meant I couldn't go into any depth on a subject. However, in the "Notes and References" section, there are suggestions for further reading and sourcing for various quotes or credentials of cited authors.

Touching on fields of study as diverse as philosophy, popular culture, computer science, sociology, and theology, the entries in this book also include topics that interested me personally. In fact, when I was younger, I became fascinated with Jasia Reichard's *Cybernetic Serendipity: The Computer and the Arts*, published in 1969; the book featured computer-generated poetry, paintings, music, graphics, and more. I'm also particularly fascinated by the strides AI experts have made in the realm of art, using generative adversarial networks (GANs) to produce amazing photorealistic images of simulated faces, flowers, and birds. GANs make use of two neural networks pitted against each other—one network generating ideas and patterns, the other judging the results.

Today, the applications of AI seem limitless, and billions of dollars are invested in the development of AI each year. As you will discover, such technologies have been used to decipher the Vatican's Secret Archives in an attempt to resolve the complicated handwritten texts in this huge historical collection. AI has also been

used for predicting earthquakes, interpreting medical images and speech, and for predicting a person's time of death based on patient information in a hospital's electronic health records. AI has been deployed for generating jokes, mathematical theorems, US patents, games and puzzles, innovative designs for antennas, new paint colors, new fragrances, and more. Today, as many of us talk to our phones and other devices, our relationship with machines will continue to become even more intimate and humanlike in the future.

Organized chronologically according to the year associated with a key event, publication, or discovery, the dating of book entries is a question of judgment. Some dates are approximate; whenever possible, I tried to give a justification for the dates used.

You'll also notice an increase in the number of entries after the year 1950. Daniel Crevier, author of *AI: The Tumultuous History of the Search for Artificial Intelligence* (1993), notes that in the 1960s, "AI blossomed in a thousand flowers. AI researchers applied their new programming techniques to many problems which, although real, had been carefully simplified, partly to isolate the problems to be addressed, but partly also to fit into the tiny memories of the computers available in those days."

The mystery of consciousness, the limits of artificial intelligence, and the nature of the mind will be studied for years to come and have actually intrigued people from ancient times. The author Pamela McCorduck suggested in her book *Machines Who Think* that AI began with an ancient wish to "forge the gods."

Future AI discoveries will be among humanity's greatest achievements. The story of AI is not only about how we will shape our future but also about how humans will mesh with a landscape of accelerating intelligence and creativity all around us. What will it mean to be "human" a hundred years from now? What will society be like, given the increasing use of AI agents? How will jobs be affected? Will we fall in love with robots?

If AI methods and models are already being used to help determine who gets hired for a job, who we date, who makes parole, who is likely to develop a psychiatric disorder, and how to autonomously drive cars and drones, then how much control over

our lives will we give to AIs of the future? As they increasingly make decisions for us, will AI units be easily *fooled* into making critical errors? How can AI researchers better understand why some machine-learning algorithms and architectures are more effective than others, while also making it easier for AI researchers to reproduce one another’s results and experiments?

Moreover, how can we ensure that AI-driven devices behave in an ethical manner, and will machines ever have mental states and feelings in the same sense that humans do? Surely, AI machines will help us think new thoughts and dream new dreams, functioning as prosthetics for our feeble brains. For me, AI cultivates a perpetual state of wonder about the limits of thought, the future of humanity, and our place in the vast space-time landscape that we call home.

Tic-tac-toe can be made more challenging for humans and AI-based machines by extending the 3 × 3 game to higher dimensions and array sizes, such as in this 4 × 4 × 4 version, and by introducing gravity effects, as each piece slides to the bottom available position.

c. 1300 BC

TIC-TAC-TOE

Archeologists can trace what appears to be “three-in-a-row games” to around 1300 BCE in ancient Egypt. For tic-tac-toe (TTT), two players, O and X, take turns marking their symbols in the spaces of a 3 × 3 grid. The player who first places three of his own marks in a horizontal, vertical, or diagonal row wins.

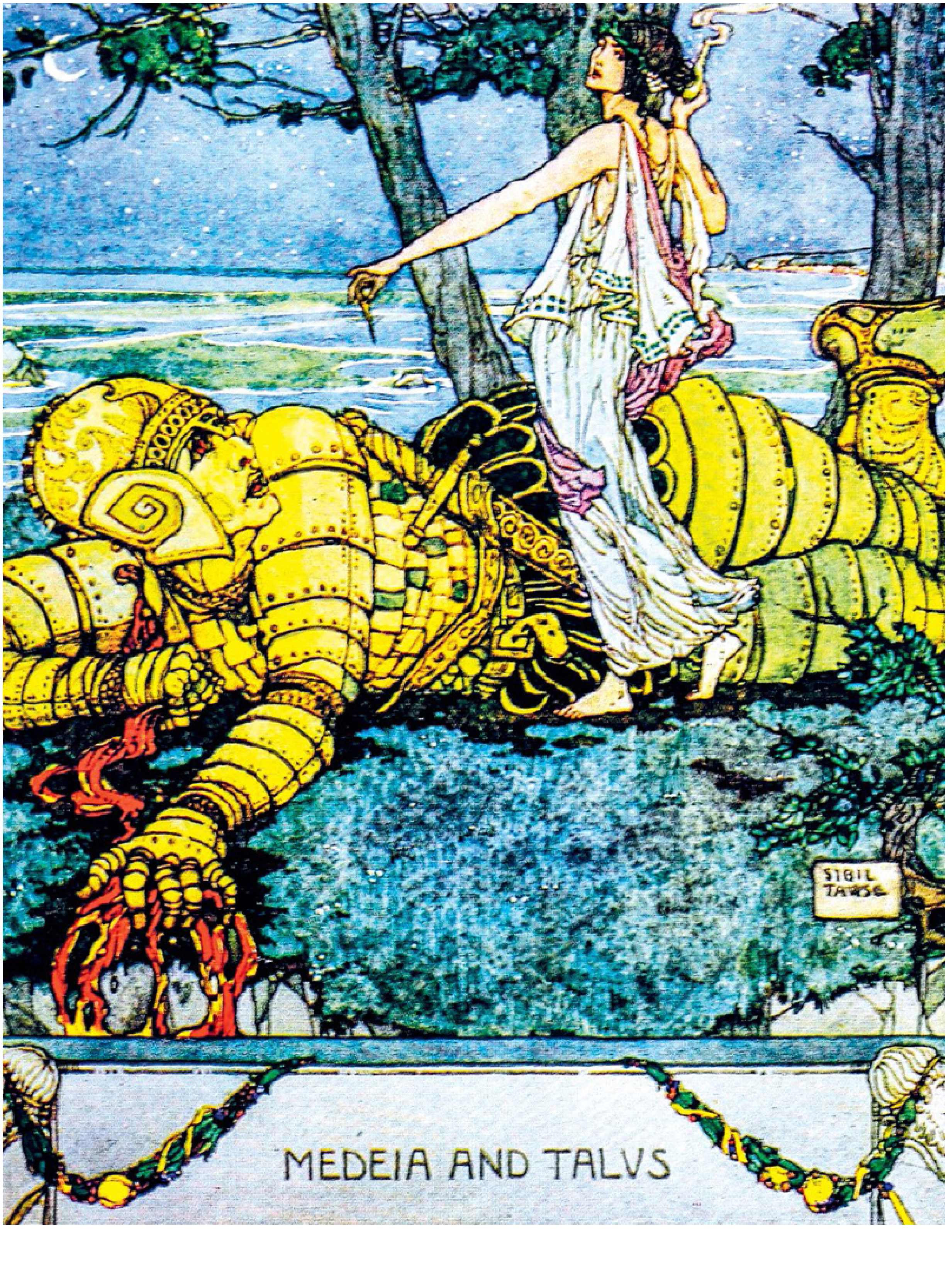
We’re interested in TTT for this book because it has often been used as an introduction to aspects of artificial intelligence and computer programming, due to the simplicity of searching its game trees (where nodes are positions in a game, and edges represent moves). TTT is a “perfect information” game, as all players know all the moves that have taken place. It is also a sequential game without randomization, with the players taking turns and avoiding the use of dice.

TTT may be considered an “atom” upon which the molecules of more advanced games of position were built through the centuries. With the slightest of variations and extensions, the simple game of TTT becomes a fantastic challenge, requiring significant time to master. Mathematicians and puzzle aficionados have extended TTT to larger boards, higher dimensions, and strange playing surfaces such as rectangular or square boards connected at their edges to form a torus (doughnut shape) or Klein bottle (a surface with just one side).

Consider some TTT curiosities. Players can place their *X*s and *O*s to fill a TTT board in 362,880 (or 9!) ways. However, there are 255,168 possible games in TTT when considering all possible games that end in 5, 6, 7, 8, and 9 moves. In 1960, the MENACE AI system (a contraption of colored beads and matchboxes) learned to play TTT using reinforcement learning. In the early 1980s, computer geniuses Danny Hillis, Brian Silverman, and friends built a Tinkertoy® computer that played TTT from 10,000 Tinkertoy pieces. In 1998, researchers and students at the University of Toronto created a robot to play three-dimensional (4 × 4 × 4) TTT with a human.

SEE ALSO The Consciousness Mill (1714), Reinforcement Learning (1951), Connect Four (1988), Othello (1997), Solving the Game of Awari (2002)

A depiction of Talos from Thomas Bulfinch's *Stories of Gods and Heroes* (1920), drawn by English artist Sybil Tawse (1886–1971).



c. 400 BCE

TALOS

“Many people are familiar with the figure of Talos,” writes author Brian Haughton, “through his depiction as a bronze giant in the 1963 movie *Jason and the Argonauts*. . . . But where did the idea for Talos come from, and could he have been the first robot in history?”

According to Greek mythology, Talos was a huge bronze automaton whose job was to protect Europa—the mother of King Minos of Crete—from invaders, pirates, and other enemies. Talos was programmed to patrol the island’s shores by circling the entirety of Crete three times each day. One way of deterring foes involved throwing huge boulders down on them. At other times, the giant robot would leap into a fire until he glowed with heat, then hug the enemy’s body to burn him to death. Talos is sometimes depicted as a winged creature, as was discovered on coins in Phaistos, Crete, from around 300 BCE; other vase paintings date to around 400 BCE.

Various explanations have been given regarding the creation and death of Talos. In one myth, he is made by Hephaestus—the Greek god of metalworking, metallurgy, fire, blacksmiths, and other artisans—at the request of Zeus. Because Talos was an automaton, his internal structure was less complex than a human’s; in fact, Talos had just one vein that ran from his neck to his ankle. The vein was sealed and protected from leaking by a bronze nail at the ankle. According to one legend, the sorceress Medea drove him mad with death spirits (Keres) and caused him to dislodge the nail. The ichor

(i.e., divine blood) then gushed out of him "like molten lead," killing him.

Talos is just one example showing how the ancient Greeks thought about robots and other automata. For another, consider the work of mathematician Archytas (428–347 BCE), who may have designed and built an automaton driven by steam in the form of a self-propelled flying bird called "The Pigeon."

SEE ALSO Ktesibios's Water Clock (c. 250 BCE), Lancelot's Copper Knights (c. 1220), Golem (1580), *Frankenstein* (1818)

This impressive bust of Aristotle is a Roman copy of a bronze original by Greek sculptor Lysippos, who lived in the fourth century BCE.

c. 350 BCE

ARISTOTLE'S *ORGANON*

The Greek philosopher Aristotle (384–322 BCE) touched upon several influential topics during his life that are still of interest to AI researchers today. In his book *Politics*, Aristotle speculated that automatons could someday replace human slaves: "There is only one condition in which we can imagine managers not needing subordinates, and masters not needing slaves. This condition would be that each instrument could do its own work, at the word of command or by intelligent anticipation, like the statues of Daedalus or the tripods made by Hephaestus, of which Homer relates that 'of their own motion they entered the conclave of Gods on Olympus,' as if a shuttle should weave of itself, and a plectrum should do its own harp playing."

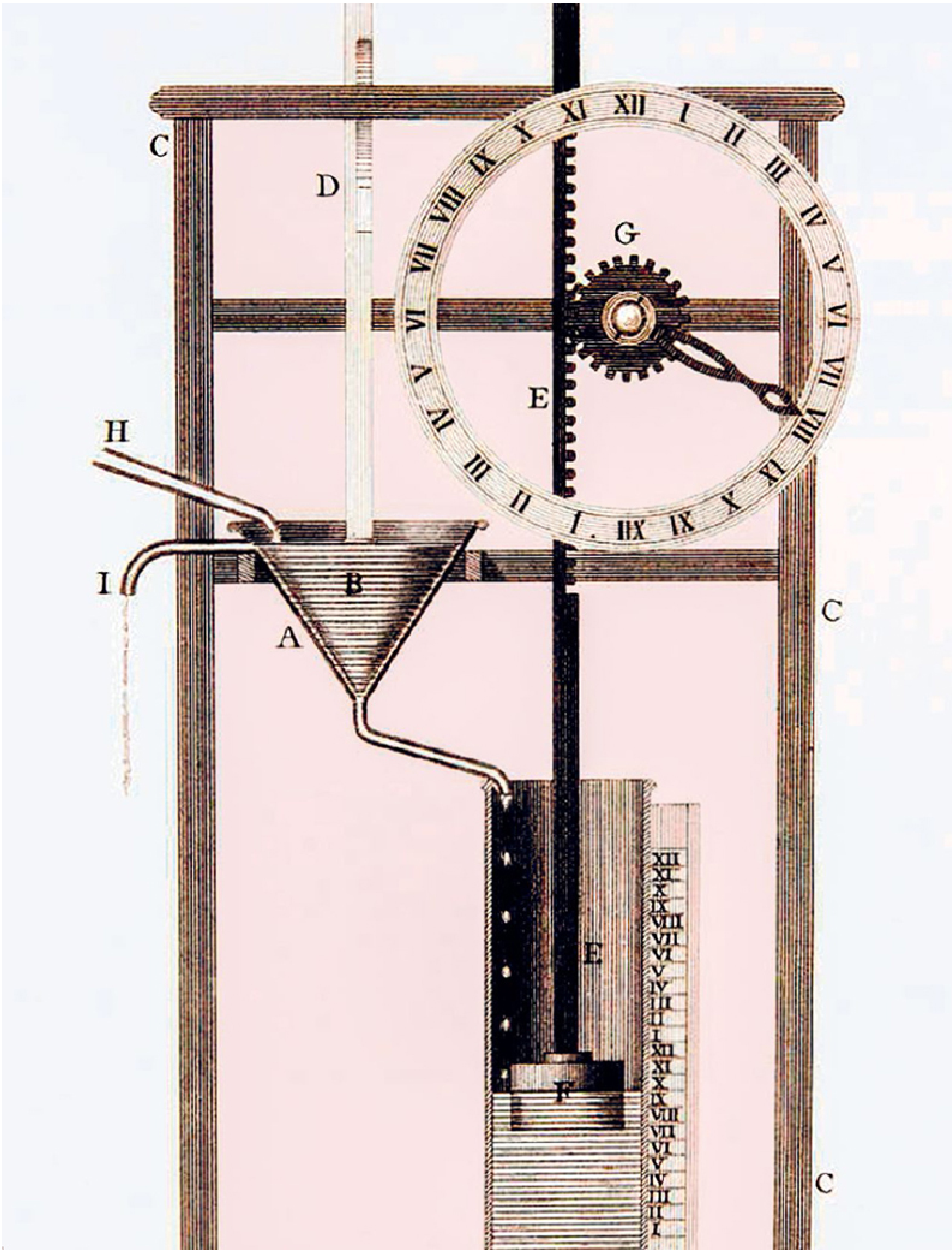
Aristotle also pioneered the systematic study of logic. In his *Organon* (Instrument) writings, he provides approaches for how to investigate truth and how to make sense of the world. The primary tool in Aristotle's toolkit is the *syllogism*, a three-step argument such as, "All women are mortal; Cleopatra is a woman; therefore Cleopatra is mortal." If the two premises are true, we know that the conclusion must be true. Aristotle also makes a distinction between particulars and universals (i.e., general categories). *Cleopatra* is a particular term, whereas *woman* and *mortal* are universal terms. When universals are used, they are preceded by the words *all*,

some, or *no*. Aristotle analyzed many possible kinds of syllogisms and showed which of them are valid.

Aristotle extended his analysis to syllogisms that involved modal logic—that is, statements containing the words *possibly* or *necessarily*. Modern mathematical logic can depart from Aristotle's methodologies or extend his work into other kinds of sentence structures, including ones that express more complex relationships and ones that involve more than one quantifier, as illustrated by this sentence: "No men like all men who dislike some men." Nevertheless, Aristotle's deep study of logic is considered to be one of humankind's greatest achievements, providing an early impetus for many developments in mathematics and AI.

SEE ALSO Talos (c. 400 BCE), Boolean Algebra (1854), Fuzzy Logic (1965)

The water clock shown here does not have all of Ktesibios's features, but it provides an illustration regarding how these devices operate. This is from *The Cyclopaedia or Universal Dictionary of Arts, Sciences and Literature* by Abraham Rees in 1820.


C
D
G
E
H
I
B
A
C
E
F
C

c. 250 BCE

KTESIBIOS'S WATER CLOCK

"Ktesibios's water clock was significant because it forever changed our understanding of what a man-made object could do," writes journalist Luke Dormehl. "Before Ktesibios's clock, only a living thing was thought to be capable of modifying its behavior according to changes in the environment. After Ktesibios's clock, self-regulating feedback control systems became a part of our technology."

The Greek inventor Ktesibios, or Tesibius (fl. 285–222 BCE), was famous in Alexandria, Egypt, for his devices involving pumps and hydraulics. His water clock, or *clepsydra* (lit. "water thief"), is of particular interest because it employed a regulator in the form of a feedback-control float that maintained a constant water-flow rate, thus allowing his timepiece to provide reasonable estimates of time according to the level of water in a receiving container. In one version of his clock, units of time are marked on a column that a humanoid figure points to as he rises with the changing water level in the reservoir. According to some reports, the humanoid figure was accompanied by other mechanisms, such as turning pillars and falling stones or eggs, along with trumpet-like sounds. Ktesibios's clepsydras were used to allocate time to speakers in court proceedings and to limit time spent by patrons in Athenian brothels.

Ktesibios was likely the first head of the Museum of Alexandria, an institution that included the Library of Alexandria and that attracted leading scholars of the Hellenistic world. Although he is famous for

his particular kinds of clepsydras, other related water clocks were also built in ancient China, India, Babylon, Egypt, Persia, and elsewhere. Ktesibios was also reported to have invented an eerie robotic statue of a deity that was featured in processions (e.g., the famous Grand Procession parade of Ptolemy Philadelphus). This automaton was able to stand up and sit down via the rotation of cams (non-circular wheels that convert circular motion to linear motion) that were perhaps linked to the movement of a cart.

SEE ALSO Al-Jazari's Automata (1206), Hesdin Mechanical Park (c. 1300), da Vinci's Robot Knight (c. 1495)

The abacus has had a huge impact on human civilization. For many centuries, this device served as a tool to allow people to perform fast calculations in commerce and in engineering. Abacuses were employed by Europeans long before they used the Hindu-Arabic numeral system.

c. 190 BCE

ABACUS

“Artificial intelligence started with the calendar and abacus,” writes engineer and author Jeff Krimmel. “Artificial intelligence is any technology that helps a human being perform a cognitive task. In this light, a calendar is a piece of artificial intelligence. It supplements or replaces our memory. Likewise, an abacus is a piece of artificial intelligence. . . . We have no need to perform complex arithmetic in our head.”

There is evidence of instruments used for performing calculations in ancient Mesopotamia and Egypt, but the oldest surviving counting board dates to around 300 BCE in Greece with the Salamis Tablet, a marble slab with several groups of parallel line markings. Other boards used in antiquity were usually wood, metal, or stone, and they contained lines or grooves along which beads or stones were moved.

Around 1000 CE, the Aztecs invented the *nepohualtzintzin* (referred to by aficionados as the “Aztec computer”), an abacus-like device that made use of corn kernels threaded through wooden frames to help operators perform computations. The modern abacus, which contains beads that move along rods, dates at least as far back as 190 CE in China, where it is called the *suanpan*. In Japan, the abacus is called the *soroban*.

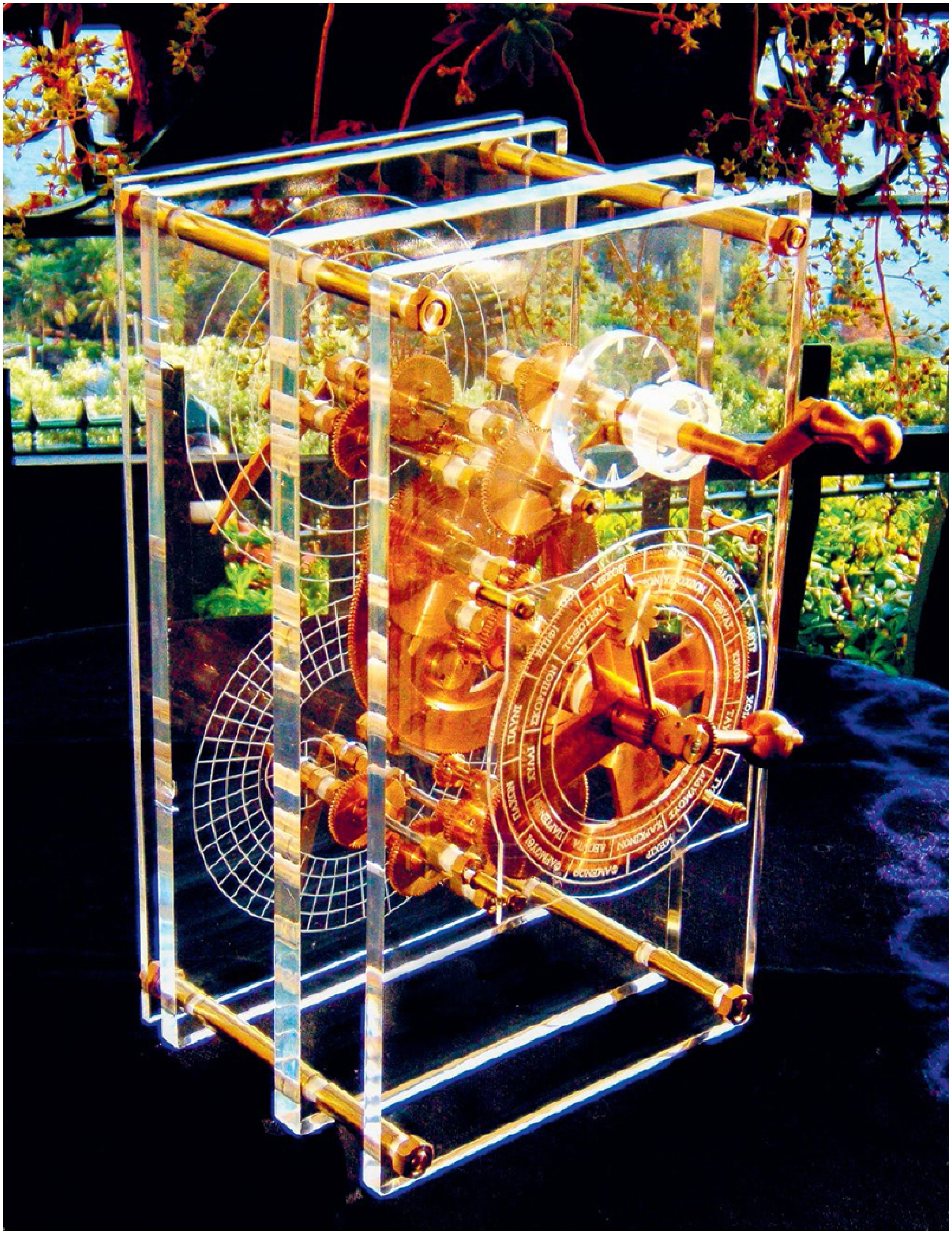
In some sense, the abacus may be considered the ancestor of the computer; and like the computer, the abacus serves as a tool to

allow humans to perform fast calculations in commerce and in engineering. With slight variations in design, abacuses are still used in China, Japan, parts of the former Soviet Union, and Africa. Although the abacus is generally used for fast addition and subtraction operations, experienced users are able to quickly multiply, divide, and calculate square roots. In 1946, a calculating competition was held in Tokyo between a Japanese soroban operator and a person using an electric calculator to see which method yielded faster results to several problems of arithmetic. In most cases, the soroban operator beat the electric calculator.

So important is the abacus that, in 2005, Forbes.com readers, editors, and a panel of experts ranked the abacus as the second most important tool of all time in terms of its impact on human civilization. First and third on the list were the knife and the compass, respectively.

SEE ALSO Antikythera Mechanism (c. 125 BCE), Babbage's Mechanical Computer (1822), ENIAC (1946)

A modern reconstruction of the Antikythera mechanism showing the gears and hand cranks.



c. 125 BCE

ANTIKYTHERA MECHANISM

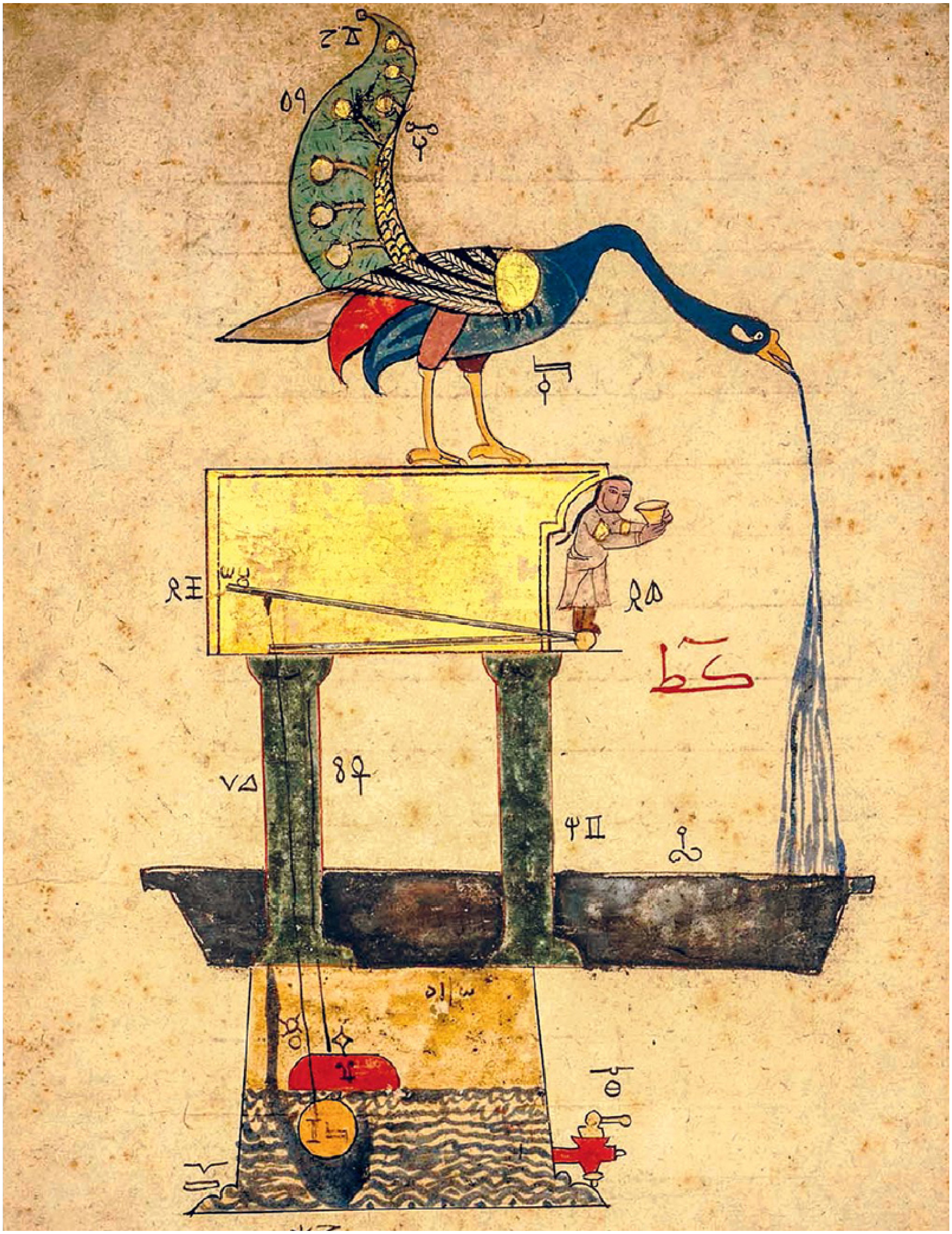
In his book *Artificial Intelligence*, psycho-logist Alan Garnham discusses the Anti-kythera mechanism and notes: “Perhaps the principal line of development that led to AI was the attempt to produce machines that took the drudgery out of human intellectual endeavor and, at the same time, eliminated some of the errors to which it is prone.” The Antikythera mechanism is an ancient geared computing device that was used to calculate astronomical positions. Discovered by sponge divers around 1900 in a shipwreck off the coast of the Greek island Antikythera, the device is thought to have been built around 150–100 BCE. As described by journalist Jo Marchant: “Among the salvaged hoard subsequently shipped to Athens was a piece of formless rock that no one noticed at first, until it cracked open, revealing bronze gearwheels, pointers, and tiny Greek inscriptions. . . . A sophisticated piece of machinery consisting of precisely cut dials, pointers, and at least thirty interlocking gear wheels, nothing close to its complexity appears again in the historical record for more than a thousand years, until the development of astronomical clocks in medieval Europe.”

A dial on the front of the device probably carried at least three hands—one indicating the date and the other two indicating the positions of the sun and the moon. The device was also probably used to track dates of ancient Olympic games, predict solar eclipses, and indicate other planetary motions.

The Moon mechanism uses a special train of bronze gears, two of which are linked with a slightly offset axis to indicate the position and phase of the moon. As is known today from Kepler's Laws of Planetary Motion, the moon travels at different speeds as it orbits Earth (e.g., faster when it is closer to Earth), and this speed differential is modeled by the Antikythera mechanism, even though the ancient Greeks were not aware of the actual elliptical shape of the orbit. Marchant writes: "By turning the handle on the box you could make time pass forwards or backwards, to see the state of the cosmos today, tomorrow, last Tuesday or a hundred years in the future. Whoever owned this device must have felt like master of the heavens."

SEE ALSO Ktesibios's Water Clock (c. 250 BCE), Abacus (c. 190 BCE), Babbage's Mechanical Computer (1822)

An intricate peacock basin from al-Jazari’s *Book of Knowledge of Ingenious Mechanical Devices* (opaque watercolor, gold, and ink on paper).



1206

AL-JAZARI’S AUTOMATA

The polymath, inventor, artist, and engineer Ismail al-Jazari (1136–1206) lived during the height of the Islamic Golden Age, following his father as the chief engineer at the Artuklu Palace in Anatolia (present-day Diyarbakır, Turkey). Al-Jazari’s *Book of Knowledge of Ingenious Mechanical Devices*—written at the request of his royal employer and published the year al-Jazari died—contains descriptions of numerous mechanical devices that al-Jazari had built, including moving human and animal automata as well as water-raising machines, fountains, and clocks. During his research and engineering, al-Jazari employed camshafts, crankshafts, escapement wheels, segmental gears, and other sophisticated mechanisms.

Among his automata are moving peacocks driven by water, a waitress that serves drinks, and a musical robot band of four automatic musicians in a boat with changing facial expressions that were controlled by rotating shafts. Some researchers have speculated that the movements of the robot band may have been programmable, indicating an extra degree of technical sophistication. His elephant clock featured a humanoid robot striking a cymbal at regular intervals, along with a robotic bird that chirped as a scribe rotated, marking out time with his pen. Al-Jazari’s 11-foot-tall (3.4 m) castle clock featured five automaton musicians.

According to English engineer and historian Donald R. Hill (1922–1994), who is famous for his English translation of al-Jazari's work, "It is impossible to over-emphasize the importance of al-Jazari's work in the history of engineering. Until modern times there is no other document from any cultural area that provides a comparable wealth of instructions for the design, manufacture, and assembly of machines. No doubt this is partly due to the fact that there was usually a social and cultural divide between those who made and those who wrote. When a scholar described a machine that had been constructed by an illiterate craftsman, he was usually interested in the finished product; he neither understood nor cared about the messy business of construction. . . . We therefore owe a great debt to [al-Jazari's employer] for our possession of a unique document."

SEE ALSO Ktesibios's Water Clock (c. 250 BCE), Hesdin Mechanical Park (c. 1300), Religious Automata (1352), Jaquet-Droz Automata (1774)

Lancelot battles the copper knight automata in order to enter the Doloreuse Garde castle. The humanoid knights are often depicted without clothing. (From *Lancelot do lac*, France, 15th century. Paris, BnF, MS Fr. 118, fol. 200v.)

c. 1220

LANCELOT'S COPPER KNIGHTS

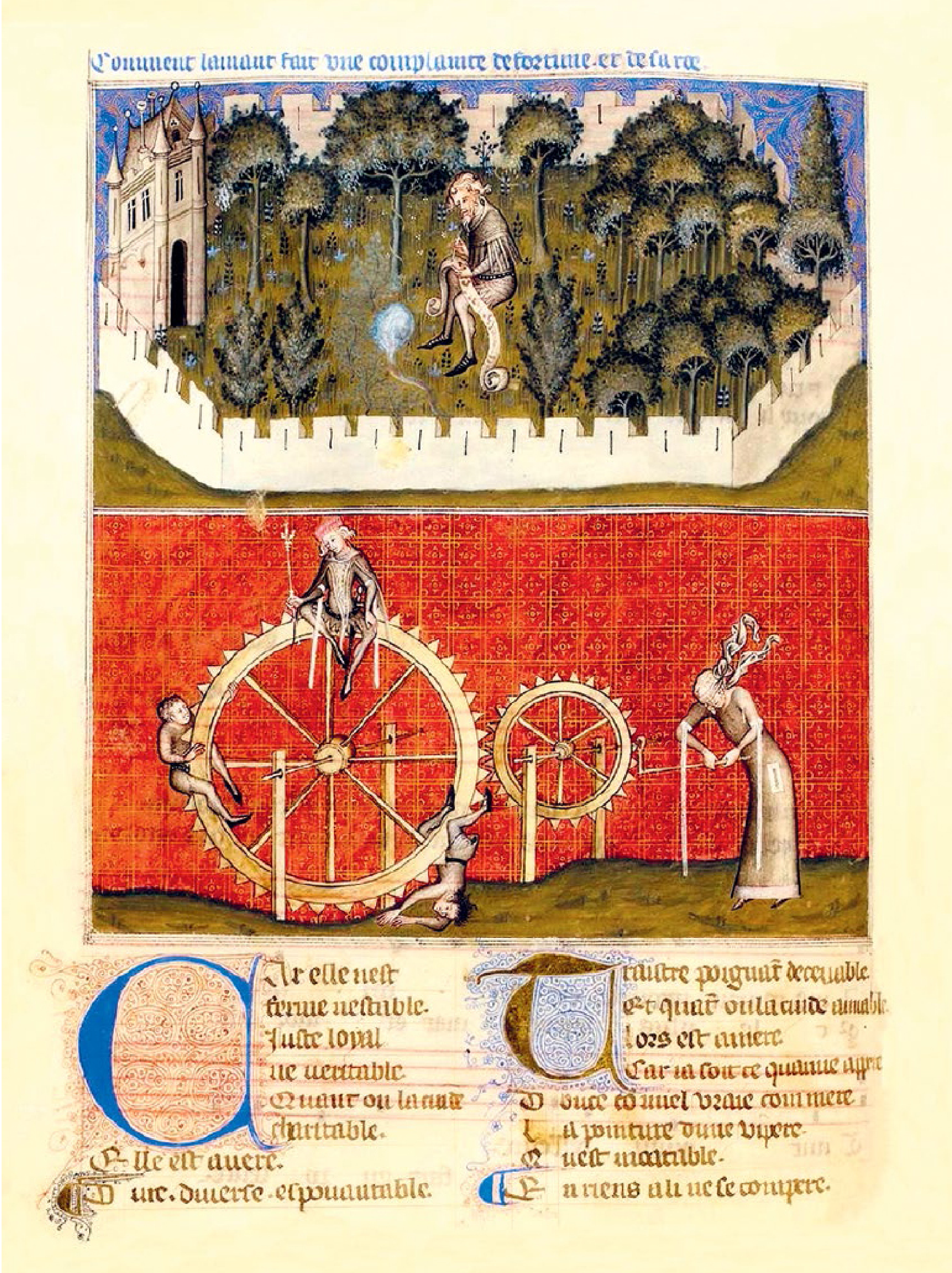
Simple examples of AI, in the form of mechanical men and creatures, became common in European medieval times, when, as historian Elly Truitt writes, "Golden birds and beasts, musical fountains, and robotic servants astound and terrify guests. . . . Automata stood at the intersection of natural knowledge (including magic) and technology, and . . . were troubling links between art and nature." Both actual and fictional devices in ancient literature provide a fascinating glimpse of an "interdependence of science, technology, and the imagination."

One famous example of fictional medieval robots occurs in *Lancelot du lac* (*Lancelot of the Lake*, c. 1220), an old French prose tale that recounts the adventures of King Arthur and his Knights of the Round Table, including the secret romance of Sir Lancelot and Arthur's wife, Guinevere. While outside Doloreuse Garde, a frightening enchanted castle, Lancelot encounters a small army of robotic copper knights. After entering the castle, he defeats two more sword-wielding copper knights as they guard an internal chamber where a young copper woman holds the keys to the enchantment. He uses the keys to open a box that contains thirty copper tubes, from which come terrifying cries—and he quickly falls asleep. Upon awakening, he finds that the copper woman has collapsed to the ground, and the copper knights are shattered.

Historian Jessica Riskin writes: “The automaton knights and damsels of Arthurian legend were accompanied by gold, silver and copper children, satyrs, archers, musicians, oracles and giants. These fictional artificial beings had plenty of real counterparts. Actual mechanical people and animals thronged the landscape of late medieval and early modern Europe.” As one example, around the same time as the tale of Lancelot’s copper knights, French artist and engineer Villard de Honnecourt (c. 1225–c. 1250) created a mechanical eagle that was designed to turn its head towards the deacon when he read from the Gospel. Riskin notes that these lifelike automata provide a context for scientific and philosophical models of living beings as machines that emerged in the seventeenth century.

SEE ALSO Talos (c. 400 BCE), Hesdin Mechanical Park (c. 1300), da Vinci’s Robot Knight (c. 1495), Golem (1580), Tik-Tok (1907)

Hesdin park. Depiction of the walled garden at Hesdin (at top). The gear image (at bottom) is the personification of Fortune, turning a mechanism that may represent the automata at Hesdin. (From Guillaume de Machaut's *Le remede de fortune,* France, c. 1350–1356. Paris BnF, MS Fr. 1586, fol. 30v.)


Comment lamant fait une complainte de fortune et de sa roe.
Car elle nest
ferme ne estable.
Juste loyal
ne veritable.
Quant on la cuide
charitable.
Elle est avere.
Dure. diverse. espouantable.
Traistre poignant decevable.
Et quant on la cuide amiable.
Lors est amere.
Douce com miel vraie commere
La pointure dune vipere.
Quest mortible.
En riens a li ne se compere.

c. 1300

HESDIN MECHANICAL PARK

Starting around 1300, the park at Hesdin, in northeastern France, evolved into a famous site for lifelike simulations of humans and animals. The Hesdin automata included androids, monkeys, birds, and a timekeeping device. The earliest Hesdin machines were built at the request of Robert II (1250–1302), Count of Artois. Examples included a bridge with six groups of mechanical monkeys covered in badger fur, to make them appear realistic. A mechanized boar head adorned the wall of a pavilion. When Robert died, his daughter Mahaut (1268–1329) became the patron of innovation and continued with the upkeep of his "engines of amusement." For example, in 1312, the monkeys were covered with new fur, and horns were added to make them appear demonic.

The idea of the automaton park may have been kindled from Islamic culture and engineers, along with automata in French romance literature. Historian Scott Lightsey writes: "Hesdin's centrality in the European idea of artificial marvels is indicative of the way this new phenomenon of wonder was replacing supernatural contingencies with the motives of court life. . . . Technological innovations allowed them to begin reenacting even the traditional supernatural themes of romance in their elaborate halls and pleasure gardens."

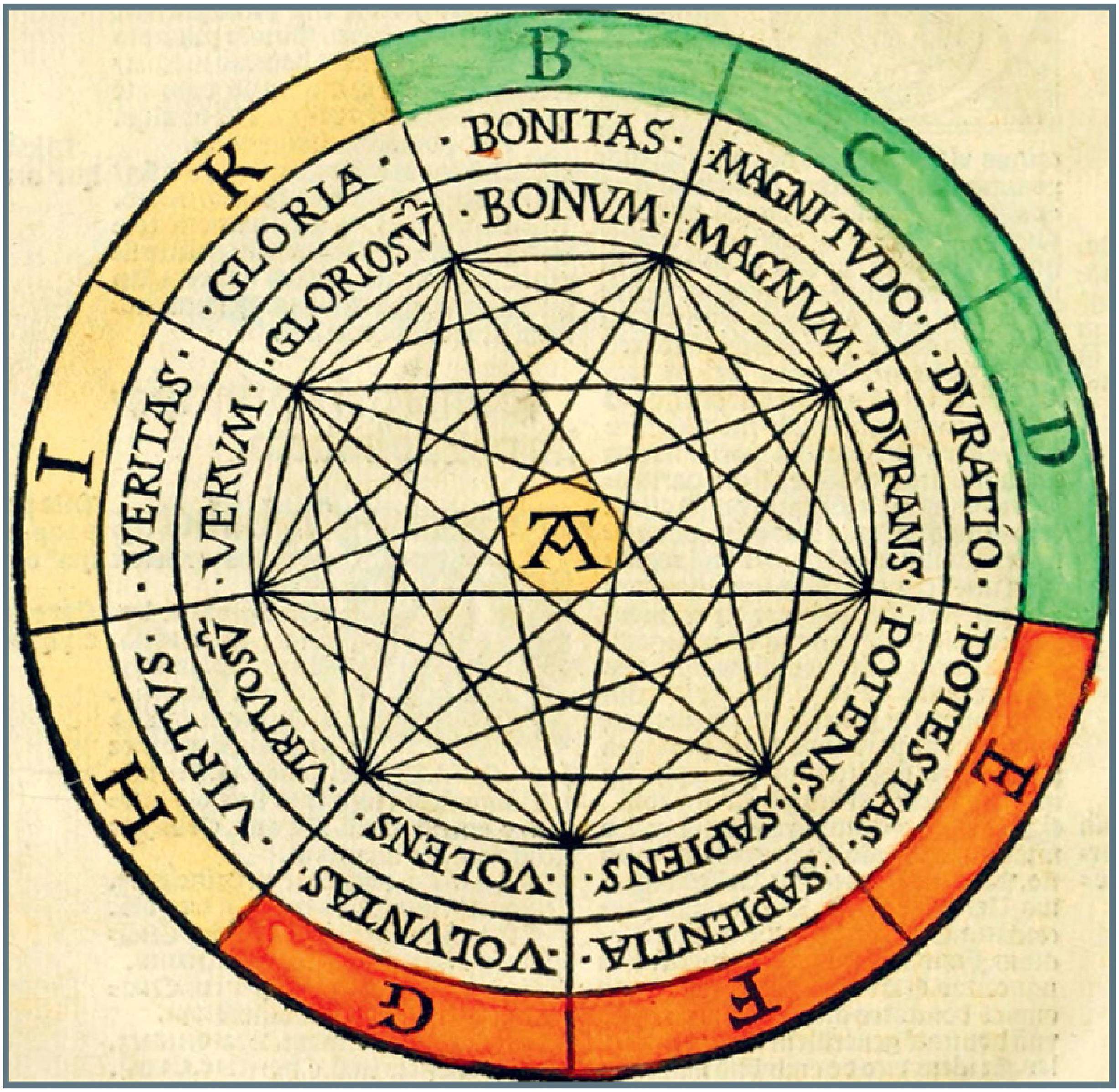
Over the years, work continued in Hesdin, with upgrades to various mechanical marvels, including a wooden hermit who spoke

to onlookers, a talking owl, and fountains with mechanical birds. The waving monkeys and other automata were most likely operated by weight-driven mechanisms, with clockwork parts and/or hydraulics.

These robotic installations offered visitors a glimpse of a future with more widespread automation. As historian Silvio A. Bedini writes: “The role of automata in the progress of technology is . . . of considerable importance. Efforts to imitate life by mechanical means resulted in the development of mechanical principles and led to the production of complex mechanisms which have fulfilled technology’s original aims—the reduction or simplification of physical labor. ”

SEE ALSO Ktesibios’s Water Clock (c. 250 BCE), Al-Jazari’s Automata (1206), Lancelot’s Copper Knights (c. 1220), Religious Automata (1352), de Vaucanson’s Duck Automaton (1738), Jaquet-Droz Automata (1774)

One set of rotating wheels and combinations of the *Ars Magna of Ramon Llull*.
(From *Illuminati sacre pagine p. fessoris amplissimi magistri Raymundi Lull*, 1517.)



c. 1305

RAMON LLULL'S *ARS MAGNA*

"The quest for artificial intelligence (AI) begins with dreams—as all quests do," writes computer scientist Nils Nilsson. "People have long imagined machines with human abilities—automata that move and devices that reason." One of the earliest devices in the history of AI is the Lullian Circle. In his book *Ars Magna* (*The Great Art*, c. 1305), Catalan philosopher Ramon Llull (c. 1232–c. 1315) included a paper construction of rotating concentric circles with letters and words written along their circumference. Much like a mechanical lock, the characters and words could be lined up in novel combinations, but in this case the combinations generated a fountain of new ideas and logical explorations. Author Martin Gardner writes: "It was the earliest attempt in the history of formal logic to employ geometrical diagrams for the purpose of discovering nonmathematical truths, and the first attempt to use a mechanical device—a primitive logic machine—to facilitate the operation of a logic system."

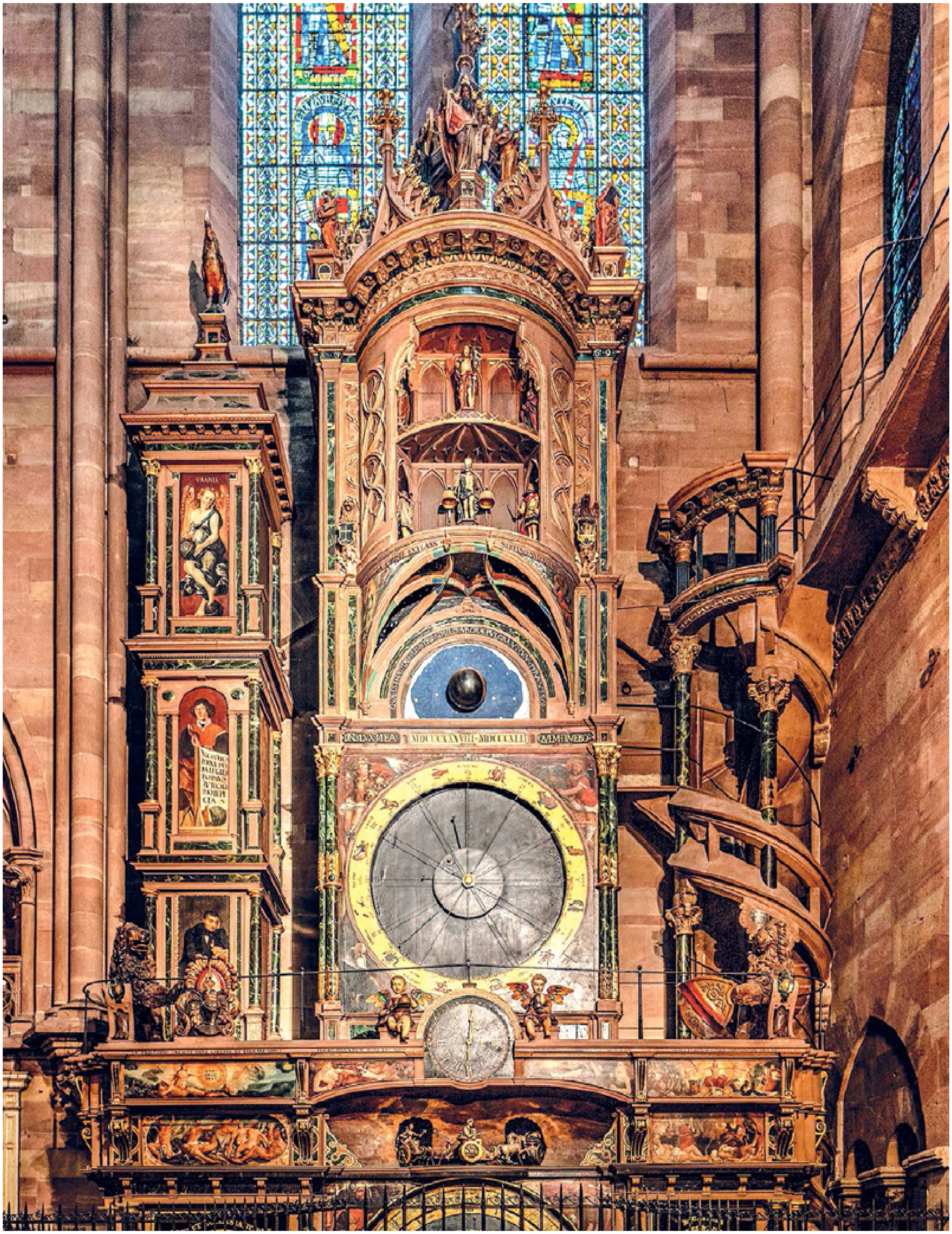
Llull's devices for combinatorial creativity provided an early method of using "logical means to produce knowledge," writes author Georgi Dalakov. "Llull demonstrated in an extremely elementary, but nevertheless workable, way that human thought can be described and even imitated by a device. This was a small step towards the thinking machine." Let us imagine Llull, sitting at a table by candlelight, turning his disks to combine words. He believed, according to author Krystina Madej, "a higher knowledge would be

revealed and would provide logical answers to questions about religion and creation. . . . [He wanted] to investigate truths and produce new proofs through these combinatory devices."

The research on formal logic by German polymath and calculus co-founder Gottfried Leibniz (1646–1716), as well as his invention of the "stepped reckoner," was stimulated by Llull's work. As data researcher and professor Jonathan Gray writes: "The initial trickles of Llull's and Leibniz's arcane combinatorial fantasies have gradually given way to ubiquitous computational technologies, practices, and ideals which are interwoven into the fabric of our worlds—the broader consequences of which are still unfolding around us . . . regardless of whether or not the machines operate in the ways that we imagine."

SEE ALSO Lagado Book-Writing Engine (1726), Computational Creativity (1821), *Cybernetic Serendipity* (1968)

The Strasbourg astronomical clock in the Cathédrale Notre-Dame of Strasbourg, Alsace, France. The automaton rooster is at the upper left.



1352

RELIGIOUS AUTOMATA

Various automata associated with the Christian Church appeared in late medieval and early modern Europe, ranging from automaton Christs to mechanical devils and Satan machines that made noises and stuck out their tongues. For example, during the fifteenth century, the "Rood of Grace" in the Boxley Abbey in Kent, England, was a mechanized likeness of Jesus on the Cross that moved his eyes, lips, and other parts of his body. Automaton angles and automaton simulations of biblical events also became common in the late fifteenth century. As recounted by history professor Jessica Riskin, "Automata were familiar features of daily life, originating in churches and cathedrals and spreading from there. Jesuit missionaries carried them to China as offerings to dramatize the power of Christian Europe."

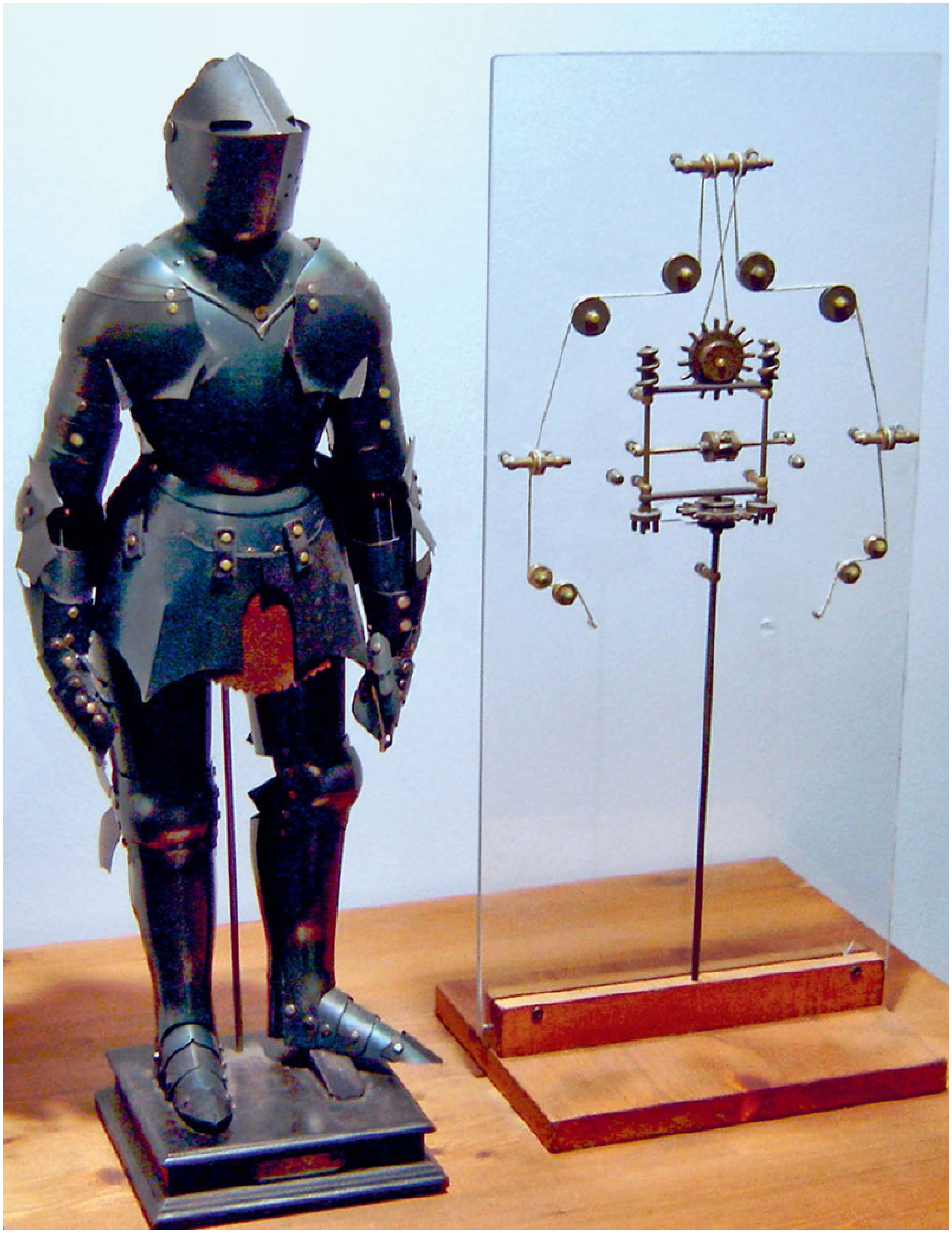
Particularly interesting is the Strasbourg astronomical clock in the Cathédrale Notre-Dame de Strasbourg in Alsace, France. Construction of the clock began in 1352, and it featured a rooster automaton that moved its head, flapped its wings, and crowed at particular times (using bellows and a reed). The clock also featured moving angels. Around 1547, this clock was replaced and upgraded, while retaining the bird automaton. The second clock stopped working in 1788. It wasn't until 1838 that the current clock came into being with new mechanisms to replace the old ones.

Aside from the automata, the Strasbourg clock features a perpetual calendar (including a means to determine the calendar date of Easter), a display of solar and lunar eclipses, and more. In 1896, author Fanny Coe wrote that the Strasbourg clock "is almost like a small theater, there are so many people and animals that have their little part to play. . . . The hours are struck by angels, and at midday and midnight life-size figures of Christ and his twelve disciples come from a door. . . . Then a gilded cock upon the upper turret of the clock flaps its wings and crows."

The scholar Julius Fraser writes: "Calendrical science and clock craftsmanship evolved to make artifacts that explained and praised the Christian universe. . . . [They were] forerunners of the later desires to put the skills of scientists and artisans to use for the earthly benefit of people."

SEE ALSO Al-Jazari's Automata (1206), Hesdin Mechanical Park (c. 1300), Jaquet-Droz Automata (1774)

Model of Leonardo's robot knight, alongside its internal gears, pulleys, and cables.



c. 1495

DA VINCI'S ROBOT KNIGHT

"Da Vinci's armored Robot Knight sat up; opened its arms and closed them, perhaps in a grabbing motion; moved its head via a flexible neck; and opened its visor," writes robotic engineer Mark Rosheim, "perhaps to reveal a frightening physiognomy. Fabricated of wood, brass, or bronze and leather, it was cable operated."

Leonardo da Vinci (1452–1519), the Italian Renaissance polymath, had interests that ranged from painting and architecture to anatomy and engineering. His journals include depictions and studies of musical instruments, crank mechanisms, and the aforementioned mechanical knight, drawn around the year 1495 in his *Codex Atlanticus* (Atlantic Codex), a twelve-volume bound set of drawings and writings. The mechanism of the Leonardo robot design —which featured articulated joints and movable arms, jaw, and head —used a pulley system for moving the robot's parts. The android wore German-Italian medieval armor, could sit up and down, and had more than one gear system that operated separately to control the upper and lower body. Leonardo sketched models for other automatons that included birds and carts.

Although we do not know if Leonardo's mechanical knight was ever built, similar robots may have inspired other engineers, such as the Italo-Spanish engineer Juanelo Turriano (c. 1500–1585), who built a mechanical monk, which used cables and pulleys, for King Philip II of Spain, who attributed his son's miraculous recovery from

a serious head injury to the divine intervention of a Franciscan missionary named Didacus. Driven by a key-wound spring, the clockwork Didacus walks while moving his mouth and arms in silent prayer. Today he is exhibited at the Smithsonian Institution in Washington, DC, and is still operational.

Reflecting on da Vinci's robot knight, authors Cynthia Phillips and Shana Priwer write "Leonardo's robot design was a culmination of his research into anatomy and geometry. What better way to combine mechanical science and human form? He took the proportions and relationships inherent in Roman architecture and applied them to the movement and life inherent in all living beings. In a way, the robot was *Vitruvian Man* brought to life."

SEE ALSO Talos (c. 400 BCE), Lancelot's Copper Knights (c. 1220), Tik-Tok (1907), Elektro the Moto-Man (1939)

The Prague Golem. In this artwork by Czech painter Eugene Ivanov, we see the golem (large central figure) and rabbi Loew ben Bezalel (perched atop the monster’s shoulder).


Cafe
33

1580

GOLEM

According to *The Forward* news-paper, “Long before Stephen Hawking warned us about the dangers of *artificial intelligence*, the legend of the *Golem* conveyed to Jews the same subliminal message.” A golem from Jewish folklore is an animated creature, crafted from clay or mud, that exhibits several forms of artificial intelligence in a man-made automaton that is difficult to control once the technology is activated and set loose in the world. Probably the most famous golem is the one that Prague rabbi Judah Loew ben Bezalel (c. 1520–1609) allegedly created in 1580 to defend the Jews of the Prague ghetto from anti-Semitic attacks. The story of this Prague golem was recorded by several authors of the 1800s.

A golem is usually inscribed with magical or religious words that keep it animated. For example, according to legend, golem creators sometimes wrote the name of God on the golem’s forehead, or on a clay tablet or paper under its tongue. Other golems were animated by the word *emet* (“truth” in the Hebrew language) written on their foreheads. By erasing the first letter to form *met* (“death” in Hebrew), the golem can be deactivated.

Some other old Jewish recipes for creating a golem required a person to combine each letter of the Hebrew alphabet with each letter from the YHVH tetragrammaton (i.e., the Hebrew name of God), and then pronounce each of the resulting letter pairs with

every possible vowel sound. The tetragrammaton serves as an "activation word" to pierce reality and energize the being.

The word *golem* appears only once in the Bible (Psalm 139:16), where it refers to an imperfect or unformed body. The New International Version translates the verse as: "Your eyes saw my unformed body. All the days ordained for me were written in your book before one of them came to be." In Hebrew, *golem* can signify a "shapeless mass" or "brainless" entity, and the Talmud uses the word to imply "imperfect." As such, most golems in literature are depicted as dumb, but they can be made to perform simple, repetitive tasks. In fact, the challenge for the golem creator was to determine how to finally stop the golem from carrying out or repeating a task.

SEE ALSO Talos (c. 400 BCE), Lancelot's Copper Knights (c. 1220), *Frankenstein* (1818)

The frontispiece of *Leviathan* by Thomas Hobbes, with an engraving by French artist Abraham Bosse (1604–1676).

1651

HOBBES'S *LEVIATHAN*

In 1651, English philosopher Thomas Hobbes (1588–1679) wrote *Leviathan*, a book that focuses on the structure of society and its relationship with government. In the book, Hobbes makes several statements that have caused science historian George Dyson to refer to Hobbes as the "patriarch of artificial intelligence." For example, in his introduction, Hobbes compares the body to a mechanical engine: "Nature (the art whereby God hath made and governs the world) is by the *art* of man . . . imitated, that [man] can make an artificial animal. For seeing life is but a motion of limbs, the beginning whereof is in some principal part within; why may we not say, that all *automata* (engines that move themselves by springs and wheels as doth a watch) have an artificial life? For what is the *heart* but a *spring*; and the *nerves* but so many *strings*; and the *joints* but so many *wheels*, giving motion to the whole body . . . ?"

When a human reasons, according to Hobbes, the person performs symbolic calculations and manipulations, something akin to addition and subtraction: "By ratiocination [thinking], I mean *computation*. Now, to compute is either to collect the sum of many things that are added together, or to know what remains when one thing is taken out of another."

Dyson asks: "If reasoning can be reduced to arithmetic, which, even in Hobbes's time could be performed by mechanism, then is a mechanism capable of reasoning? Can machines think?" Computer

architect Daniel Hillis offers his own speculation on the possibility of creating an artificial mind that can think for itself: “For those who fear mechanistic explanations of the human mind, our ignorance of how local interactions produce emergent behavior offers a reassuring fog in which to hide the soul. Although individual computers and individual computer programs are developing the elements of artificial intelligence, it is in the larger networks (or the network at large) that we are developing a more likely medium for the emergence of the Leviathan of artificial mind.”

SEE ALSO The Consciousness Mill (1714), “The Artist of the Beautiful” (1844), “Darwin among the Machines” (1863), *Giant Brains, or Machines That Think* (1949), *The Human Use of Human Beings* (1950)

If we believe that consciousness is the result of patterns and dynamic interrelationships of neurons and other cells in the brain, then our thoughts, emotions, and memories might be replicated in the movements of twigs and leaves or in the flocking of birds. (Watercolor painting by Benjavisa Ruangvaree.)

1714

THE CONSCIOUSNESS MILL

If we believe that consciousness is the result of patterns and dynamic inter-relationships of brain cells and their components, then our thoughts, emotions, and memories might be replicated in moving assemblies of Tinkertoys®. The Tinkertoy minds would need to be very big to represent the complexity of our minds, but perhaps a very complex mechanism could be created, in the same way researchers have made tic-tac-toe playing computers out of 10,000 Tinkertoys. In principle, our minds could be hypostatized in the movements of leaves and twigs or in the flocking of birds. In 1714, the German philosopher and mathematician Gottfried Leibniz in his treatise *The Monadology* imagined an AI-infused machine as big as a mill that was capable of thinking and feeling. He also realized that if we could explore inside, we would find "nothing but pieces that push one against the other and never anything to account for a perception." Similarly, it's possible that conscious AI entities could be developed in the future, even if they are not formed from wet organic matter.

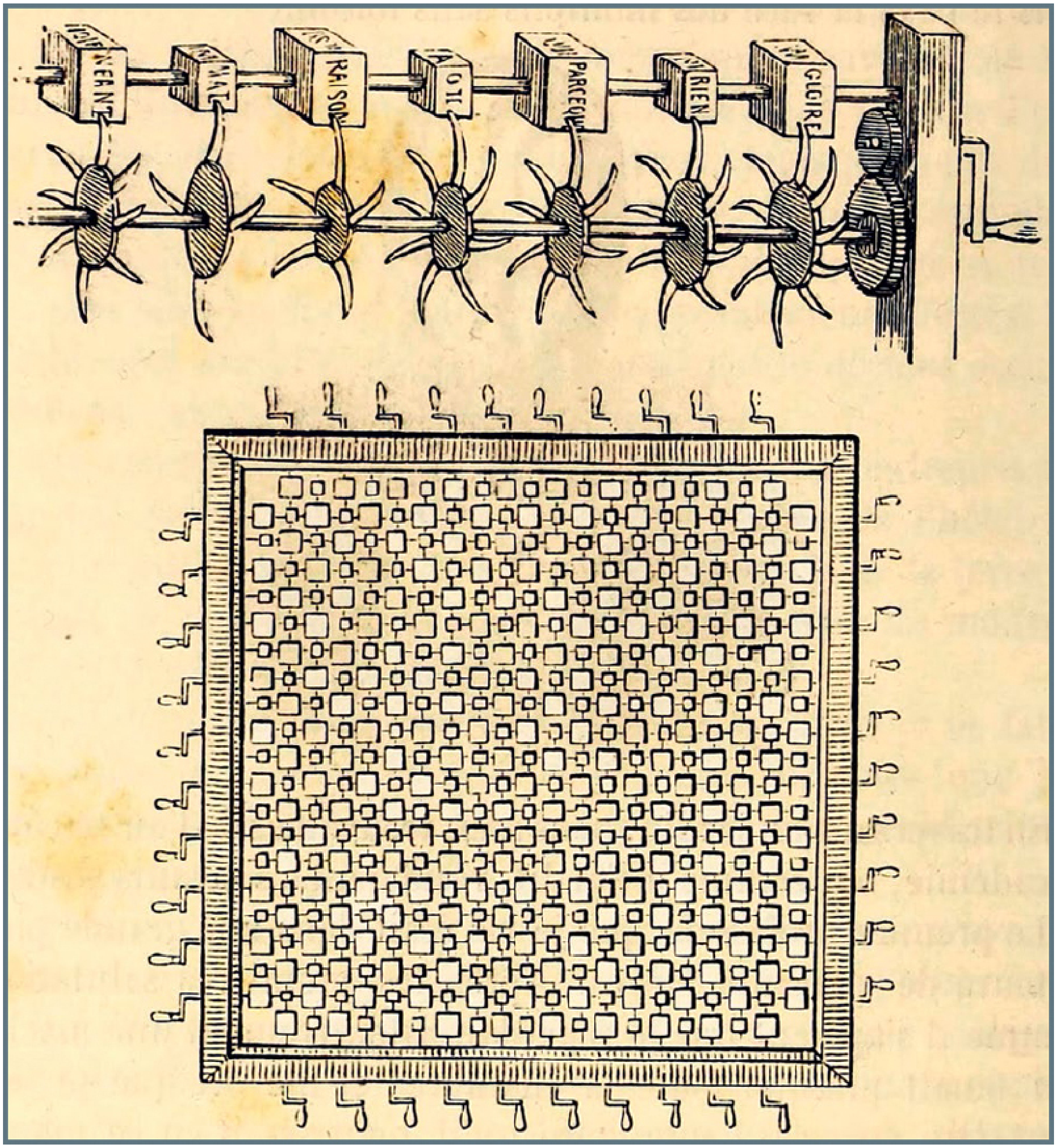
Philosopher Nick Bostrom (b. 1973) considers a single electronic brain cell: "The brain cell is a physical object with certain characteristics. If we come to fully understand these characteristics and learn to replicate them electronically, then surely our electronic brain cell can perform the same function as an organic one. And if it can be done with a brain cell, why wouldn't the resulting system be just as conscious as a brain?"

As roboticist Hans Moravec (b. 1948) writes: “We’re a conscious being simulated on a bunch of neural hardware, and the conscious being is only found in an interpretation of things that go on in the neural hardware. It’s not the actual chemical signals that are squirting around, it’s a certain high-level interpretation of an aggregate of those signals, the only thing that makes consciousness different from other interpretations, like the value of a dollar bill.”

Similarly, perhaps your brain could function even if partitioned into a hundred little boxes, separated by great distances and connected by wires or optical fibers. To understand this better, imagine your left and right brain hemispheres a mile apart and connected by an artificial corpus callosum. You would still be you, right?

SEE ALSO Tic-Tac-Toe (c. 1300 BCE), Hobbes’s *Leviathan* (1651), de Vaucanson’s Duck Automaton (1738), Searches for the Soul (1907), *Giant Brains, or Machines That Think* (1949), Living in a Simulation (1967)

The writing machine of Lagado, illustrated by French artist J. J. Grandville (1803–1847), which appeared in a French translation of *Gulliver's Travels* from 1856.



1726

LAGADO BOOK-WRITING ENGINE

Gulliver's Travels, a popular novel published in 1726 by Anglo-Irish author Jonathan Swift (1667–1745), describes a mechanical creativity engine that may be the first AI device extensively discussed in a novel. While Gulliver is in the fictional city of Lagado, a professor shows him a device for generating literature, technical books, and interesting ideas. Gulliver explains how, "By his contrivance, the most ignorant person, at a reasonable charge, and with a little bodily labor, might write books in philosophy, poetry, politics, laws, mathematics, and theology, without the least assistance from genius or study."

Gulliver is led to the device, which occupies twenty square feet in area and has various wooden surfaces "linked together by slender wires." The tile-like surfaces are covered with paper on which "were written all the words of their language, in their several moods, tenses, and declensions; but without any order."

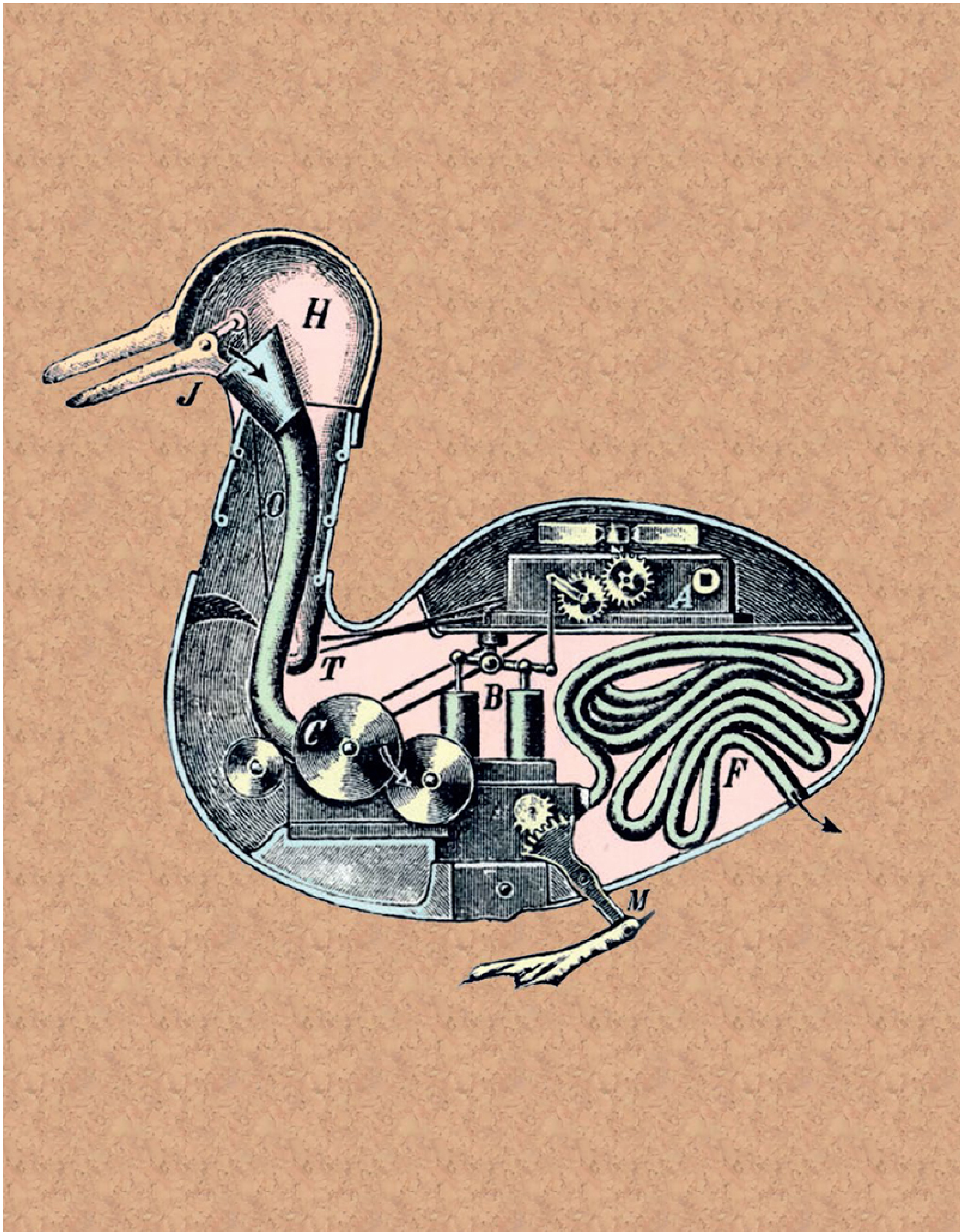
Gulliver described the device's operation: "The pupils, at his command, took each of them hold of an iron handle, whereof there were forty fixed round the edges of the frame; and giving them a sudden turn, the whole disposition of the words was entirely changed. He then commanded [the pupils] to read the several lines softly, as they appeared upon the frame; and where they found three or four words together that might make part of a sentence, they dictated to the four remaining boys, who were scribes. . . . At every

turn, . . . the words shifted into new places, as the square bits of wood moved upside down.”

Author Eric A. Weiss writes: “[The machine’s] purpose, the claims of its illustrious professor-inventor, his call for public funding, and the operation of the device by students clearly classify it as an early attempt at artificial intelligence and have caused it to be cited often as typical of this discipline.” Later forms of combinatorial, random, or artificial creativity in the real world include Racter, a computer program that generated prose for the book *The Policeman’s Beard Is Half Constructed*, published in 1984.

SEE ALSO Ramon Llull’s *Ars Magna* (c. 1305), Computational Creativity (1821), *Cybernetic Serendipity* (1968)

Popular illustration of de Vaucanson's duck automaton, which appeared in the January 21, 1899 issue of *Scientific American*. Although the mechanism portrayed here does not closely resemble the actual internal structure, the arrow is nicely placed to indicate the exit route.


H
J
O
T
A
B
C
F
M

1738

DE VAUCANSON'S DUCK AUTOMATON

"In 1738, the 29-year-old French watchmaker Jacques de Vaucanson [1709–1782] exhibited in the garden of the Tuileries what may be one of the most celebrated robots of all time," writes American neuroscientist Paul Glimcher. The de Vaucanson duck had hundreds of moving parts and feathers. It moved its head, muddled water with its bill, flapped its wings, quacked, gulped food from the exhibitor's hand, and carried out many more realistic actions. After a few minutes, remains of digested food would be excreted below. Of course, the duck was not really digesting food, and the tail end of the duck was secretly preloaded with simulated excrement. Nevertheless, such a versatile automaton triggered discussions about the line between the living and the purely mechanical, as well as the degree to which such boundaries might be blurred as robotic entities became even more versatile.

Through time, fascination with the famous digesting duck has grown, and the odd creature even makes an appearance in Thomas Pynchon's highly acclaimed 1997 novel *Mason & Dickson*, where the creature becomes conscious and follows and terrorizes a French chef with its *bec de la mort* ("beak of death").

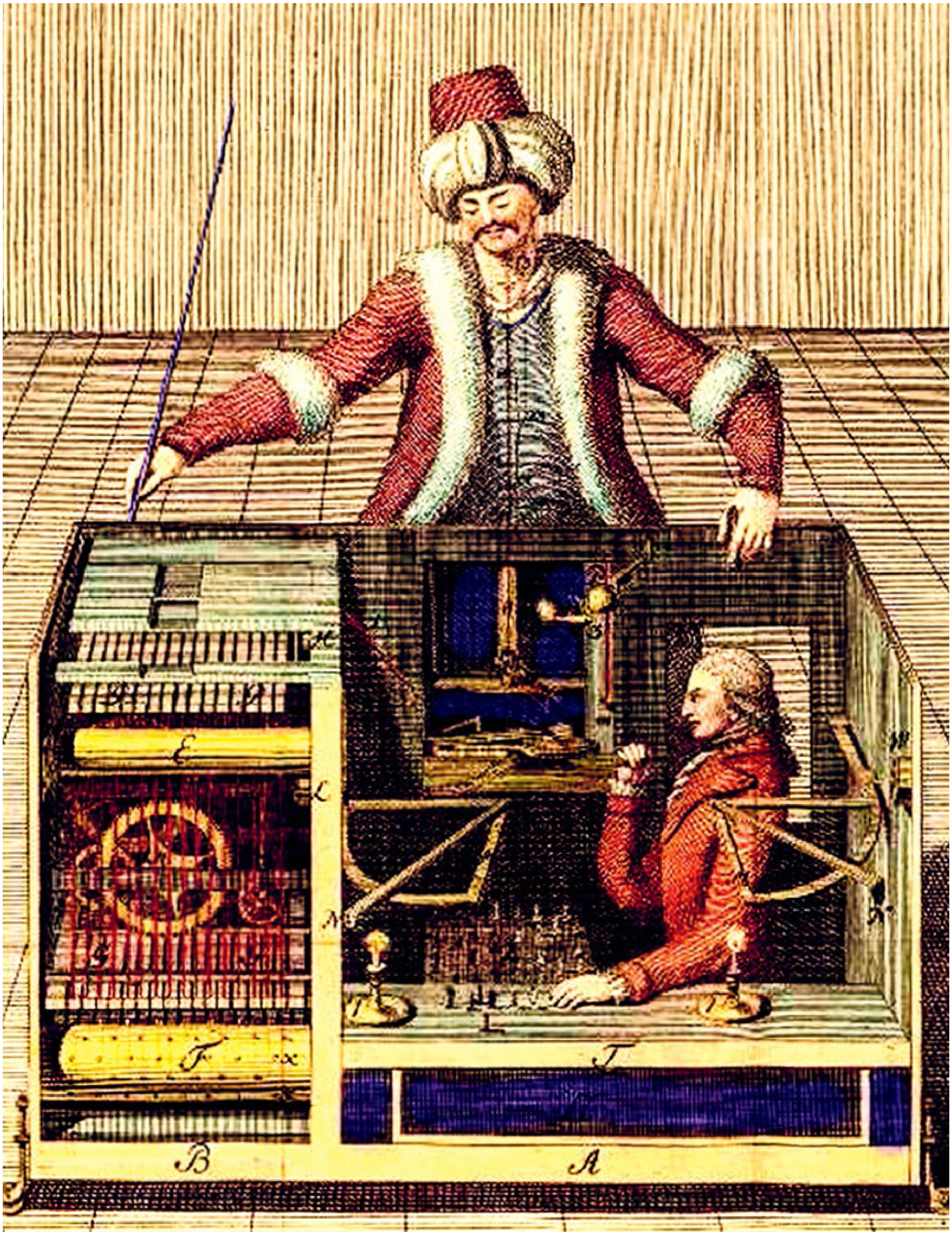
De Vaucanson also created a marvelous automaton flute player, driven by several bellows attached to three windpipes. Gears and cams triggered levers that controlled the flute-player's fingers,

tongue, and lips. The mechanized flautist "was the first example of what Diderot's *Encyclopédie* defines as an androïde, that is, a human figure performing human functions," writes historian Jessica Riskin. More immediately practical, in the 1740s de Vaucanson designed machines to weave silk—which, alas, caused (human) silk workers to rebel and pelt him with stones in the street.

As Glimcher writes: "Vaucanson's duck, raised for eighteenth-century audiences ancient questions that still haunt modern neuroscience: Are the mechanical interactions that occur inside each of us sufficient to generate the complex patterns of behavior that we actually produce? What is it that defines us as human beings, the complexity of the behavior that we produce or the specific patterns of interacting matter that appear to generate our behavior?"

SEE ALSO Hesdin Mechanical Park (c. 1300), da Vinci's Robot Knight (c. 1495), The Consciousness Mill (1714), *The Steam Man of the Prairies* (1868), *Electric Bob's Big Black Ostrich* (1893)

A peek inside the Turk, from Joseph von Racknitz (1744–1818), hypothesizing how the Turk operated. (From *Über den schachspieler des herrn von Kempelen und dessen nachbildung* [*About the Chess-Player of Mr. von Kempelen and its Replica*], Leipzig und Dresden 1789.)


E
F
T
B
A

1770

MECHANICAL TURK

The Mechanical Turk was a chess-playing android created in 1770 by Hungarian inventor Wolfgang von Kempelen (1734–1804) and presented to the Habsburg empress Maria Theresa of Austria. The machine appeared to play an excellent game of chess, as it defeated players in Europe and the Americas, including such notables as Napoleon Bonaparte and Benjamin Franklin. The life-sized android, adorned with a robe, turban, and black beard, sat at a large cabinet with a chessboard at top and actually used its hand to move chess pieces. The secret of its operation was not well known for many years, but today we understand that the complex cabinet cleverly concealed a human chess expert who used magnets to move the pieces and various levers to move parts of the android. To deepen the mystery, Kempelen would actually open the cabinet doors before play to reveal clockwork machinery inside and apparently no visible space for a human to hide. Even if many people understood the Turk to be a sophisticated “trick,” it nevertheless caused people to wonder what kinds of work machines were capable of—and what human capabilities machines might replace.

Many inaccurate articles were written about how the Turk could have operated. For example, Edgar Allan Poe suggested, incorrectly, that the player sat inside the Turk android’s body. Interestingly, Charles Babbage, one of the fathers of the modern computer, was likely inspired by the Turk, as Babbage wondered about whether

machines *could* “think,” or at least perform highly sophisticated computations, when he began to work on his mechanical computing machines.

Author Ella Morton notes: “Though the [Turk] ultimately relied on human behavior and a bit of old-fashioned magic, its convincingly mechanical nature was cause for both wonder and concern. Arriving smack-bang in the middle of the industrial revolution, the Turk raised unsettling questions about the nature of automation and the possibility of creating machines that could think. The fact that the Turk appeared to operate on clockwork mechanisms . . . contradicted the idea that chess was . . . ‘the province of intellect alone’.”

SEE ALSO Babbage’s Mechanical Computer (1822), “Elephants Don’t Play Chess” (1990), Checkers and AI (1994), Deep Blue Defeats Chess Champion (1997)

The Jaquet-Droz writer android, at the Musée d’Art et d’Histoire (Museum of Art and History) in Neuchâtel, Switzerland.

1774

JAQUET-DROZ AUTOMATA

“They might have been huge marionettes,” writes novelist Jean Lorrain (1845–1906), “or tall mannequin dolls left behind in panic—for I divined that some plague . . . had swept through the town and emptied it of its inhabitants. I was alone with these simulacra of love . . . obsessed by the fixed and varnished eyes of all those automata.”

Such eerie thoughts of lifelike automata remind us of our long fascination with robot-like beings and of a particular set of eighteenth-century automata that might serve as examples of early ancestors of computers, given the complexity and programmability of these androids. Created by watchmaker Pierre Jaquet-Droz (1721–1790) between 1768 and 1774, these three automata—the boy writer android (made from approximately 6,000 parts), the female musician (2,500 parts), and the child draughtsman (2,000 parts)—drew large crowds of admirers. The boy android dipped his writing quill in ink and could be programmed with a series of cams to write messages of up to forty characters in length. He periodically re-inked the pen, and his eyes followed his words as he wrote.

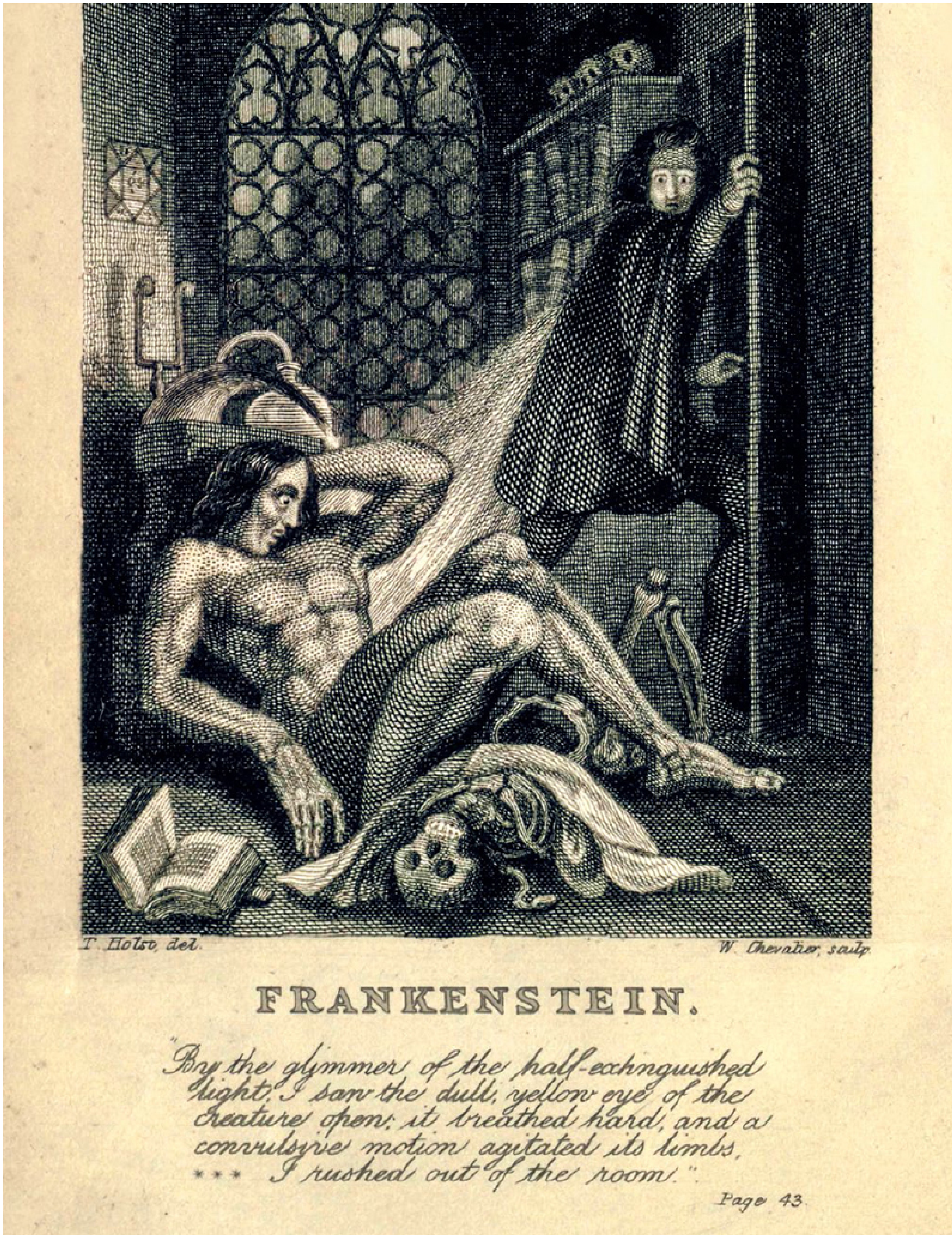
The musician automaton played an organ by actually pressing the keys with her fingers. She seemed to come alive, making natural body and head motions and following her fingers with her eyes. She was designed to continue “breathing” before and after her performances, and her body heaved with seeming emotion in time with the music. The draughtsman could sketch four different figures:

a dog, a portrait of Louis XV, cupid driving a chariot, and a royal couple.

These automata are notable in that their mechanisms reside inside their bodies (and not, for example, in a nearby piece of furniture), making the programmability, miniaturization, and haunting lifelike quality all the more impressive. Jaquet-Droz was helped by his son Henri-Louis and mechanic (and adopted son) Jean-Frédéric Leschot, and he is said to have later constructed two artificial hands for a man with congenital deformities. The hands were clothed in white gloves and were reportedly versatile enough to enable the user to write and draw.

SEE ALSO al-Jazari's Automata (1206), Hesdin Mechanical Park (c. 1300), Religious Automata (1352), de Vaucanson's Duck Automaton (1738)

Frontispiece from the 1831 edition of *Frankenstein,* published in London by Colburn and Bentley.



1818

FRANKENSTEIN

"F*rankenstein* was written during the first Industrial Revolution," writes Paolo Gallo, the Chief Human Resources Officer at World Economic Forum, "a period of enormous changes that provoked confusion and anxiety for many. It asked searching questions about man's relationship with technology: Are we creating a monster we cannot control, are we losing our humanity, our compassion, our ability to feel empathy and emotions?"

The dangers of a special kind of artificial intelligence is a prominent theme in the novel *Frankenstein; or, The Modern Prometheus* (1818) by Mary Shelley (1797–1851). In the novel, scientist Victor Frankenstein robs slaughterhouses and cemeteries in order to construct a creature from various parts, which he then animates with "a spark of life." Meanwhile, he reflects on his creation as an experiment in achieving immortality: "Life and death appeared to me ideal bounds, which I should first break through, and pour a torrent of light into our dark world. A new species would bless me as its creator. . . . I thought, that if I could bestow animation upon lifeless matter, I might . . . renew life where death had apparently devoted the body to corruption."

By the time Mary Shelley finished her novel at age nineteen, Europeans were fascinated by theories about the role of electricity in biology and the potential for reanimation of dead tissue. The basic story idea came to her in a dream. As it happens, Italian physicist

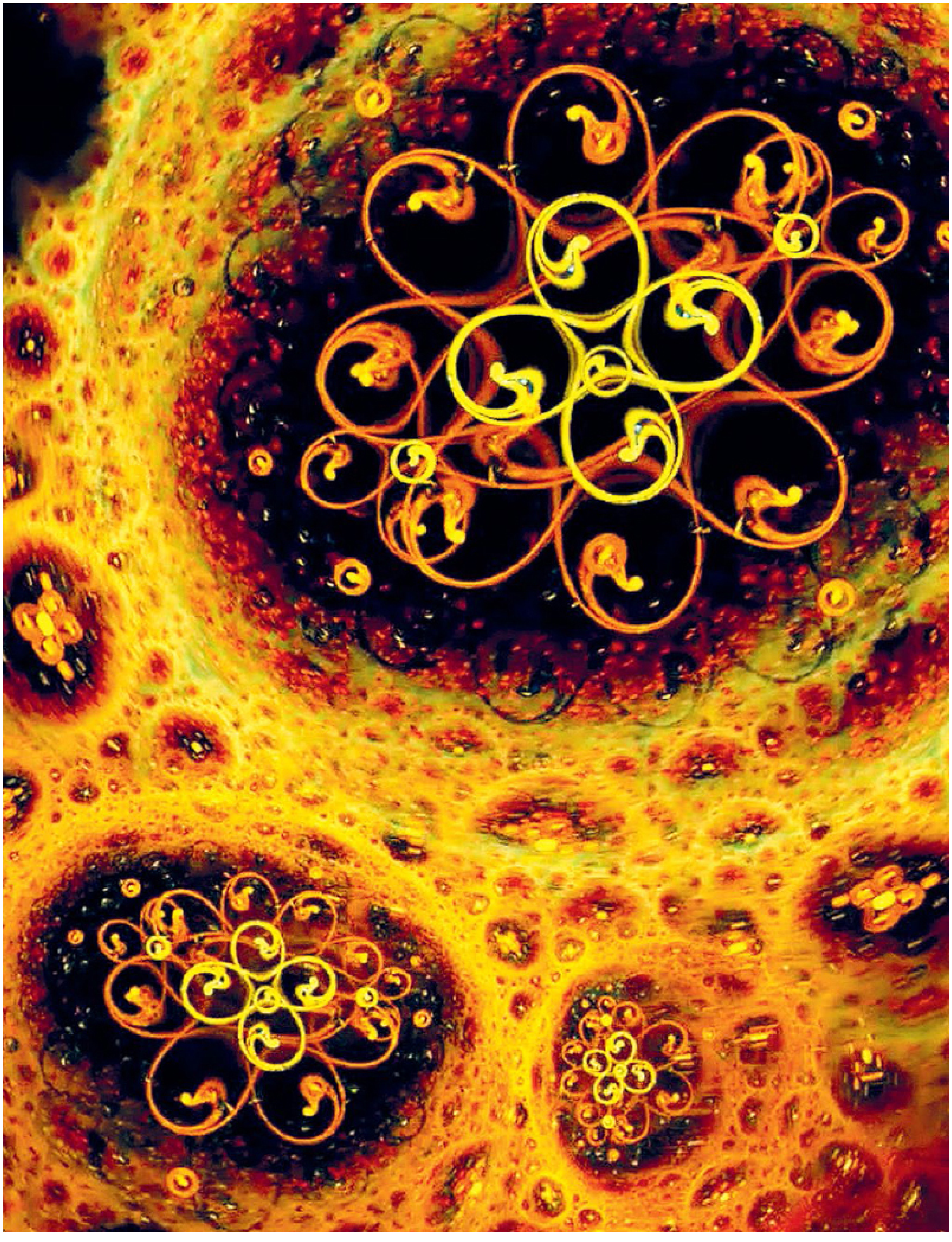
Giovanni Aldini (1762–1834) had participated in many public attempts at human reanimation through electricity around 1803 in London.

Death and destruction lurk everywhere in the novel, as numerous characters meet their ends. Notably, Victor destroys the unfinished female companion for his monster, and the monster (who is never actually referred to by the name “Frankenstein”) kills Elizabeth, the scientist’s wife. By the end of the novel, Victor has pursued his creation to the North Pole, where Victor dies, and the creature vows to kill himself on his own funeral pyre.

Journalist Daniel D’Addario notes: “*Frankenstein* relies on the notion that humans will inherently reject artificial intelligence as unnatural and bizarre. A great deal of that is owed to the particularly odd appearance of Frankenstein’s monster. . . . But what about when A.I. comes in a more attractive package, one that has real utility?”

SEE ALSO Talos (c. 400 BCE), Golem (1580), *The Steam Man of the Prairies* (1868), Rossum’s Universal Robots (1920)

Electric Sheep artwork. “Electric Sheep” refers to a collaborative abstract artwork system developed by Scott Draves. The more popular sheep live longer and reproduce according to a genetic algorithm with mutation and crossover.



1821

COMPUTATIONAL CREATIVITY

"As a society, we are jealous of our creativity," write Simon Colton and Geraint Wiggins of the Computational Group at Goldsmith College. "Creative people and their contributions to cultural progression are highly valued. Moreover, creative behavior in people draws on a full set of intelligent abilities, so simulating such behavior represents a serious technical challenge for Artificial Intelligence research. As such, we believe it is fair to characterize Computational Creativity as a frontier for AI research beyond all others—maybe, even, the final frontier."

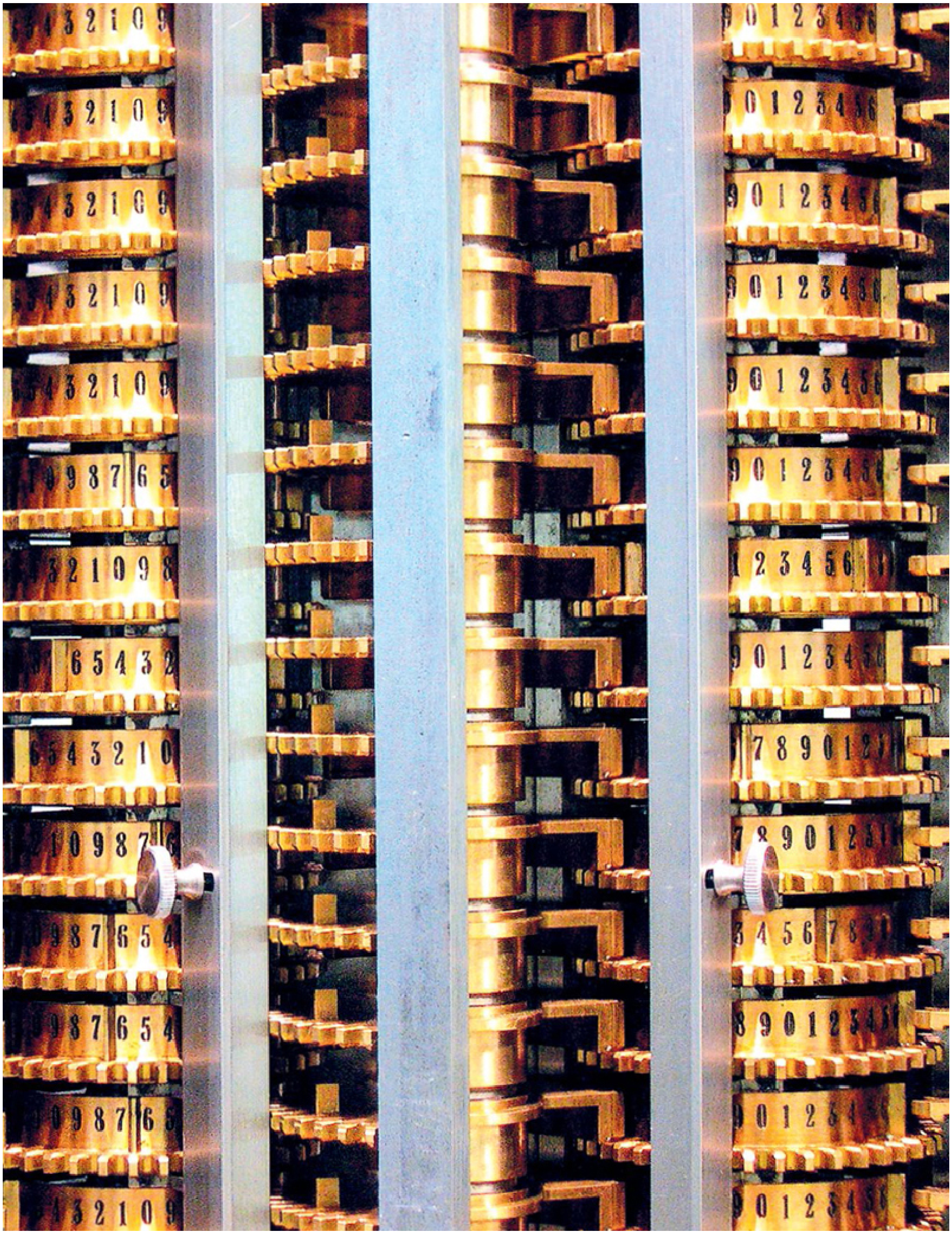
"Computational creativity," or CC, has several meanings. Here, we refer to a subfield of AI that simulates creativity using a computer or other machine. The results often seem to be novel and potentially useful. "CC" can also refer to programs that enhance human creativity. For example, using artificial neural networks (ANNs) and other methods, researchers have produced beautiful music or paintings in the style of previous artists. In particular, generative adversarial networks (GANs) may employ two contesting ANNs to produce photorealistic images of simulated faces, flowers, birds, and room interiors. Other CC approaches are used for formulating tasty cooking recipes, producing new kinds of visual art, creating poetry and stories, and for generating jokes, mathematical theorems, US patents, new games, novel chess puzzles, innovative designs for antennas and heat exchangers, and more. In short, by some form of

computational or artificial means, a computer can generate designs and output that, if any human produced them, would be deemed to be a creative act.

For an early, simple example of CC, consider that in 1821 Dietrich Winkel (1777–1826) invented the *componium*, an automatic mechanical musical organ that could produce seemingly endless variations on a musical theme. It had two barrels that took turns performing two measures of randomly chosen music. A flywheel acted as a kind of programmer to determine the selections. Winkel said that if each performance lasted five minutes, the componium would require more than 138 trillion years to produce every possible combination of music!

SEE ALSO Ramon Llull's *Ars Magna* (c. 1305), Lagado Book-Writing Engine (1726), *Cybernetic Serendipity* (1968), Genetic Algorithms (1975), Computer Art and DeepDream (2015)

Working model of a portion of Charles Babbage's Difference Engine, currently located at the London Science Museum.



1822

BABBAGE'S MECHANICAL COMPUTER

Charles Babbage (1791–1871) was an English analyst, statistician, and inventor who, in 1819, saw the Mechanical Turk, as this mechanical man toured England and defeated human chess players. Of course, Babbage must have known that the Turk was some kind of trick, but many have suggested that the android device *inspired* Babbage to ponder other kinds of more practical thinking machines, as an early step on the journey toward artificial intelligence.

Babbage is often considered the most important mathematician-engineer involved in the "prehistory" of computers. In particular, he is famous for conceiving an enormous hand-cranked mechanical calculator, an early progenitor of our modern computers. Babbage thought the device would be most useful in producing mathematical tables, but he worried about mistakes that would be made by humans who transcribed the results from its thirty-one metal output wheels. Today we realize that Babbage was around a century ahead of his time and that the politics and technology of his era were inadequate for his lofty dreams.

Babbage's Difference Engine, begun in 1822 but never completed, was designed to compute values of polynomial functions, using about 25,000 mechanical parts. He also had plans to create a more general-purpose computer, the Analytical Engine, which could be programmed using punch cards and had separate areas for number

storage and computation. Estimates suggest that an Analytical Engine capable of storing one thousand 50-digit numbers would be over one hundred feet in length. Ada Lovelace (1815–1852), the daughter of the English poet Lord Byron, gave specifications for a program for the Analytical Engine. Although Babbage provided assistance to Ada, many consider Ada to be the first computer programmer.

In 1990, novelists William Gibson and Bruce Sterling wrote *The Difference Engine*, which asked readers to imagine the consequences of Babbage's mechanical computers becoming available to Victorian society. In fact, at the end of the novel, we encounter an alternate, fictional 1991, in which a self-aware computer has evolved and appears to be the narrator of the book.

SEE ALSO Abacus (c. 190 BCE), Mechanical Turk (1770), ENIAC (1946)

“The Artist of the Beautiful” describes the creation of a delicate and beautiful butterfly automaton with mystical and lifelike qualities.

1844

"THE ARTIST OF THE BEAUTIFUL"

"The Artist of the Beautiful" by Nathaniel Hawthorne (1804–1864) is the first robotic-insect short story, and a personal favorite for both its haunting beauty and questions it raises about AI and humanity's response. Published in 1844, well before the invention of the electric lightbulb, the story centers on the life of genius Owen Warland, who works in a watch shop. Owen is a sensitive young man who is secretly in love with Annie Hovenden, the shopkeeper's daughter, and wonders if it is possible to "imitate the beautiful movements of Nature, as exemplified in the flight of birds or the activity of little animals."

Owen eventually succeeds in making a mechanical butterfly. The shopkeeper discovers an early model and accidentally almost crushes the "mechanical something, as delicate and minute as the system of a butterfly's anatomy." He screams "Owen! there is witchcraft in these little chains, and wheels, and paddles."

In the closing scene of the story, Owen decides to show Annie a new version he has created: "A butterfly fluttered forth, and, alighting on her finger's tip, sat waving the ample magnificence of its purple and gold-speckled wings, as if in prelude to a flight. It is impossible to express by words the glory, the splendor, the delicate gorgeousness, which were softened into the beauty of this object. Nature's ideal butterfly was here realized in all its perfection; not in the pattern of such faded insects as flit among earthly flowers, but of

those which hover across the meads of Paradise, for child-angels and the spirits of departed infants to disport themselves with."

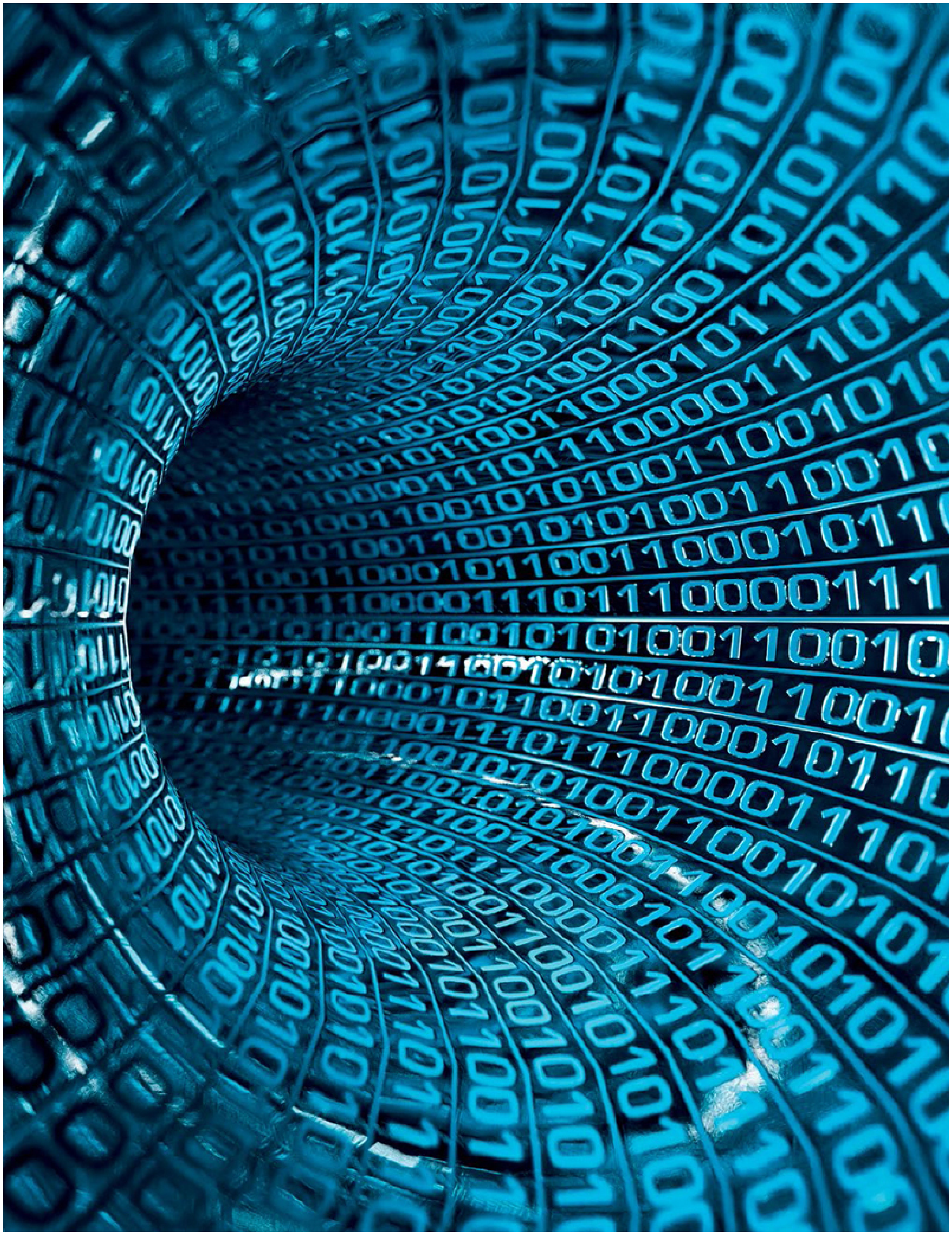
"'Beautiful! Beautiful!' exclaimed Annie. 'Is it alive? Is it alive?'"

The insect flies into the air and flutters around Annie's head. The story appears to have a sad ending for the butterfly, as it is crushed by an uncaring child into a "small heap of glittering fragments," but Owen has an epiphany of sorts, realizing that the butterfly's beauty is eternal.

As a curious aside, in 2015, US patent 9,046,884 was granted for robotic butterflies that sense, respond, and move according to the emotional states of humans to help improve our moods.

SEE ALSO de Vaucanson's Duck Automaton (1738), Jaquet-Droz Automata (1774), Tik-Tok (1907), Spielberg's *A.I. Artificial Intelligence* (2001)

When considering Boolean algebra, George Boole wrote that one of his purposes was to “investigate the fundamental laws of those operations of the mind by which reasoning is performed.”



1854

BOOLEAN ALGEBRA

On the 200th anniversary of the birth of English mathematician George Boole (1815–1864), journalist James Titcomb celebrated him as “an early thinker on the theory of artificial intelligence, believing that all human thought could be reduced to a series of mathematical rules, and advocating machinery as a way to replace human drudgery.”

In his most important work, Boole wrote that his purpose was to “investigate the fundamental laws of those operations of the mind by which reasoning is performed and . . . to collect . . . some probable intimations concerning the nature and constitution of the human mind.” His influential 1854 book was titled *An Investigation into the Laws of Thought, on Which Are Founded the Mathematical Theories of Logic and Probabilities*. Boole was interested in reducing logic to a simple algebra involving just two quantities, 0 and 1, and three basic operations: *and*, *or*, and *not*. In modern times, Boolean algebra has had vast applications in telephone switching and the design of modern computers.

In his landmark book, Boole also wrote that his goal was to “unfold the secret laws and relations of those high faculties of thought by which all beyond the merely perceptive knowledge of the world and of ourselves is attained or matured.” Fellow British mathematician Augustus De Morgan (1806–1871), who coined the term *mathematical induction*, praised Boole’s work in his posthumously

published book *A Budget of Paradoxes*: "Boole's system of logic is but one of many proofs of genius and patience combined. . . . That the symbolic processes of algebra, invented as tools of numerical calculation, should be competent to express every act of thought, and to furnish the grammar and dictionary of an all-containing system of logic, would not have been believed until it was proved."

Approximately seventy years after Boole's death, American mathematician Claude E. Shannon (1916–2001) was introduced to Boolean algebra while still a student, and he showed how Boolean algebra could be used to optimize the design of systems of telephone routing switches. He also demonstrated that circuits with relays could solve Boolean algebra problems. Thus, Boole, with Shannon's help, provided one of the foundations for our Digital Age.

SEE ALSO Aristotle's *Organon* (c. 350 BCE), Abacus (c. 190 BCE), Fuzzy Logic (1965)

In "Darwin among the Machines," Samuel Butler wrote "We are ourselves creating our own successors. . . . In the course of ages, we shall find ourselves the inferior race."

1863

“DARWIN AMONG THE MACHINES”

English author and polymath Samuel Butler (1835–1902) provided early insights into possible AIs of the future, presaging concepts such as self-improving machine superintelligence and its potential risk. In his astounding 1863 essay “Darwin among the Machines,” Butler discusses the future of “mechanical life”: “We are ourselves creating our own successors; we are daily adding to the beauty and delicacy of their physical organization; we are daily giving them greater power and supplying by all sorts of ingenious contrivances that self-regulating, self-acting power which will be to them what intellect has been to the human race. In the course of ages, we shall find ourselves the inferior race.”

With startling perceptivity, Butler envisions the gradual takeover of human endeavor by machines: “We are becoming more subservient to them; more men are daily bound down as slaves to tend them . . . devoting the energies of their whole lives to the development of mechanical life. . . . The time will come when the machines will hold the real supremacy over the world and its inhabitants. . . .”

In *The Book of the Machines* (1872), Butler muses that a mollusk does not appear to have much consciousness, yet human consciousness evolved. Similarly, machines will develop consciousness, and he asks us to “reflect upon the extraordinary advance which machines have made during the last few hundred years, and note how slowly the animal and vegetable kingdoms are

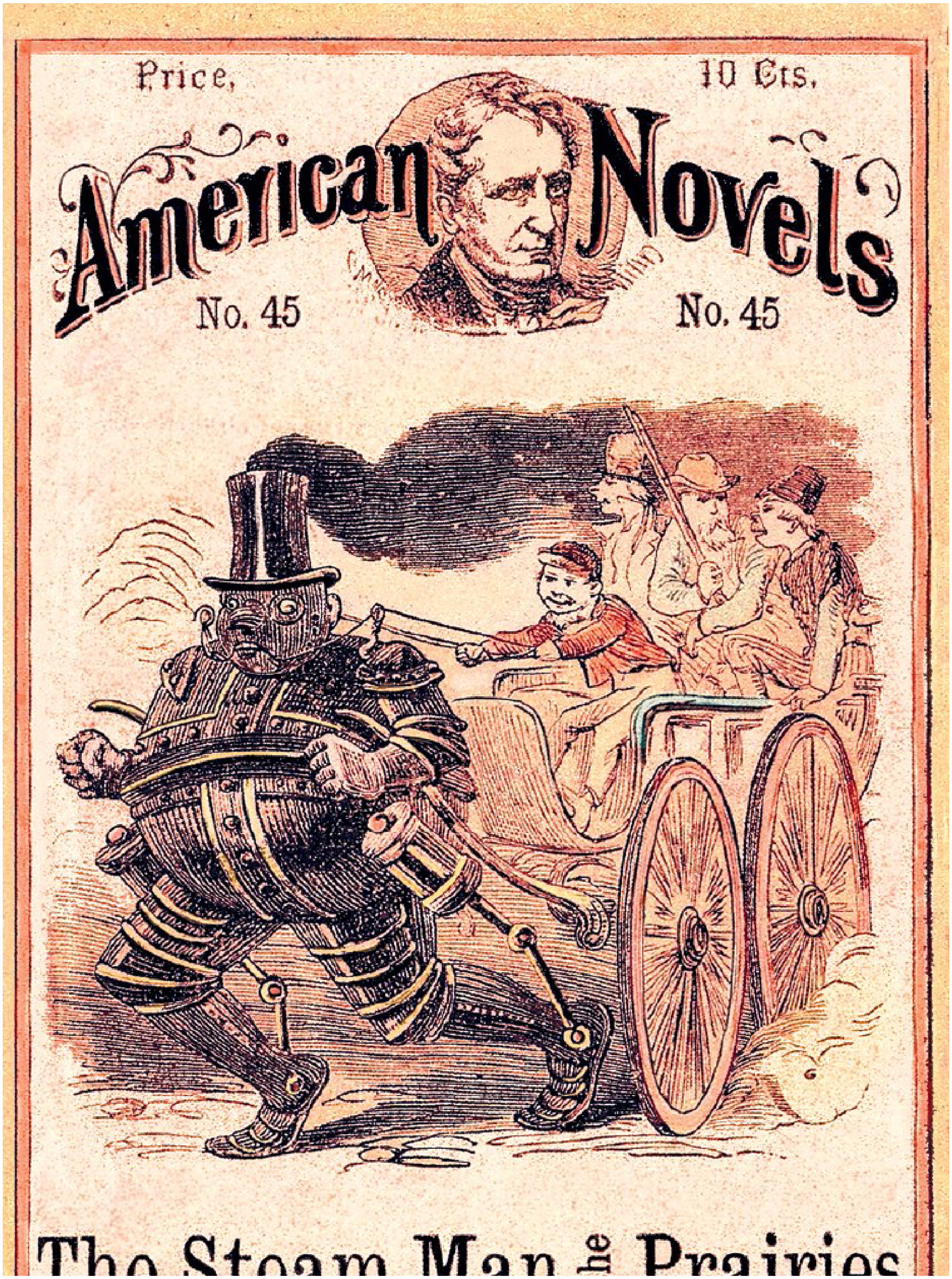
advancing. The more highly organized machines are creatures not so much of yesterday, as of the last five minutes. . . .”

Butler’s ideas echoed into the twentieth century when Norbert Wiener (1894–1964), father of the field of cybernetics, wrote: “If we move in the direction of making machines which learn and whose behavior is modified by experience, we must face the fact that every degree of independence we give the machine is a degree of possible defiance of our wishes. The genie in the bottle will not willingly go back in the bottle, nor have we any reason to expect them to be well disposed to us.”

With twenty-first-century technology now interweaved in all aspects of human life, both Butler’s and Wiener’s musings on AI appear particularly prophetic.

SEE ALSO Hobbes’s *Leviathan* (1651), *The Human Use of Human Beings* (1950), Intelligence Explosion (1965), Leakproof “AI Box” (1993), Paperclip Maximizer Catastrophe (2003)

Cover of *The Steam Man of the Prairies* by Edward S. Ellis (1868).



The Steam Man of the Prairies

FOR SALE BY

"The American News Co.,"

119 & 121 Nassau Street, N. Y.

1868

THE STEAM MAN OF THE PRAIRIES

One of the first depictions of a mechanical human in US “dime novels” (inexpensive paperbacks) occurs in *The Steam Man of the Prairies* by Ohio-born author Edward S. Ellis (1840–1916), which was reissued numerous times between 1868 and 1904. In Ellis’s novel, a teenage inventor named Johnny Brainerd creates a ten-foot-tall robot and takes it around the American Midwest. The steam man wears a stovepipe hat, pulls a wagon, and walks or runs on carefully tooled legs as fast as sixty miles per hour. Johnny, his friends, and his robot have adventures on the prairie chasing buffalo, frightening Indians, and assisting in gold mining.

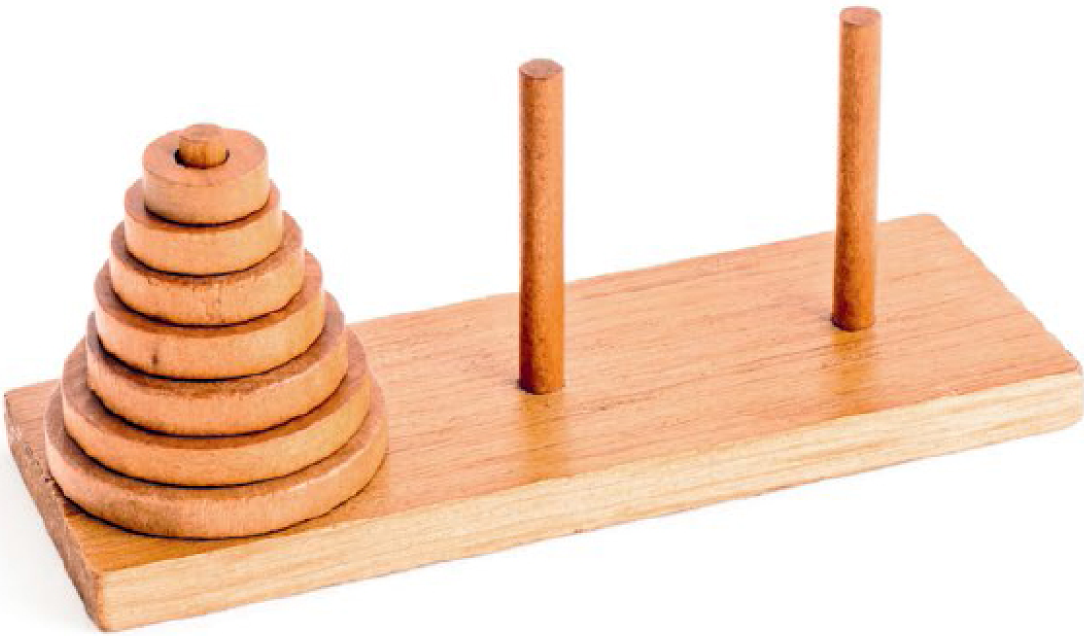
Ellis describes the steam man as “exceedingly corpulent, swelling out to aldermanic proportions, which, after all, was little out of harmony with its immense height.” The robot was fat so as to provide sufficient room for all of its machinery. Furthermore, the “face was made of iron . . . with a pair of fearful eyes, and a tremendous grinning mouth. . . . The step was natural, except when running, at which time, the bolt uprightness in the figure showed different from a human being.”

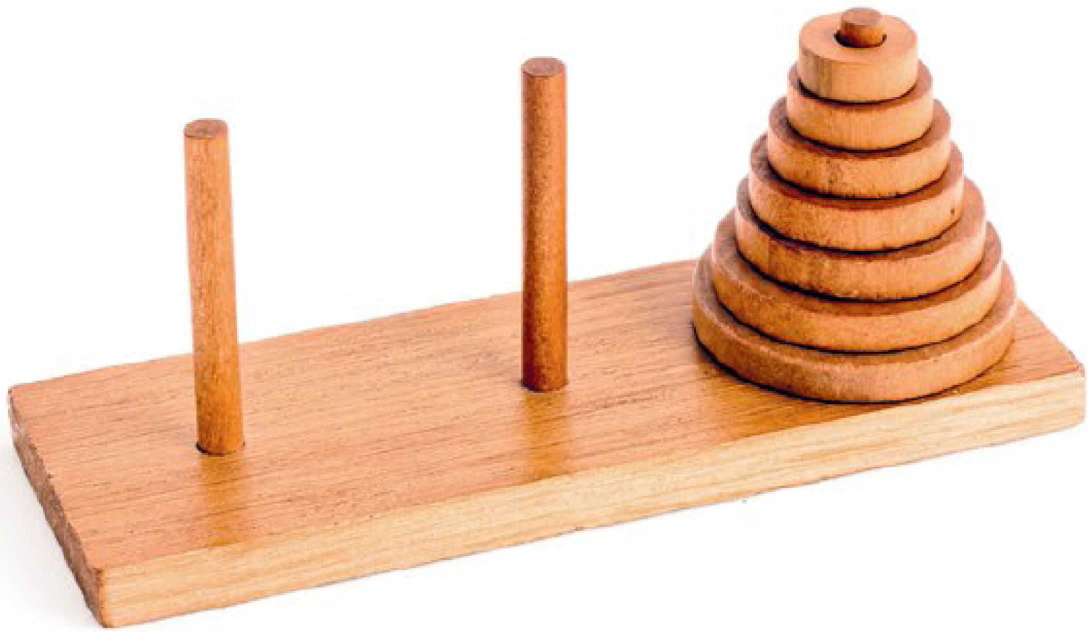
The Steam Man was inspired by an actual steam-powered humanlike robot, patented in 1868 by American inventors Zadoc P. Dederick and Isaac Grass. The fascinating Ellis series is also an early example of the “Edisonade” genre of stories that generally featured a young male inventor who uses his ingenuity to escape

from dangers. As noted by historian Andrew Liptak: “Ellis, in writing about adventures on the American frontier, was exploring the lives of people living on the edge of the unknown, much like modern stories set far out in the solar system. . . . Ellis’s works serve as a good window into the attitudes of the time, and provide some context for how the future was viewed, as well as how much the world was changing.”

SEE ALSO *Electric Bob’s Big Black Ostrich* (1893), Tik-Tok (1907), Elektro the Moto-Man (1939)

In the Tower of Hanoi game, a player can move one disk at a time to another peg by removing the top disk in any stack and placing it on the top of any other stack. A disk cannot be placed on top of a smaller disk.





1883

TOWER OF HANOI

The Tower of Hanoi has intrigued the world since it was invented by French mathematician Édouard Lucas (1842–1891) in 1883 and sold as a toy. This mathematical puzzle consists of several disks of different sizes that slide onto any of three pegs. The disks are initially stacked on one peg in order of size, with the smallest disk at the top. When playing the game, one can move one disk at a time to another peg by removing the top disk in any stack and placing it on the top of any other stack. A disk cannot be placed on top of a smaller disk. The goal is to move the entire starting stack (usually with seven or eight disks) to another peg. The minimum number of moves turns out to be $2^n - 1$, where n is the number of disks.

The original game was said to be inspired by a legendary Indian temple in which Brahmin priests continually moved 64 golden disks, using the same rules as in the Tower of Hanoi. According to this legend, when the last move of the puzzle is completed, the world will end. Note that if the priests were able to move disks at the rate of one per second, then $2^{64} - 1$, or 18,446,744,073,709,551,615, moves would require roughly 585 billion years—about 42 times the current estimated age of our universe.

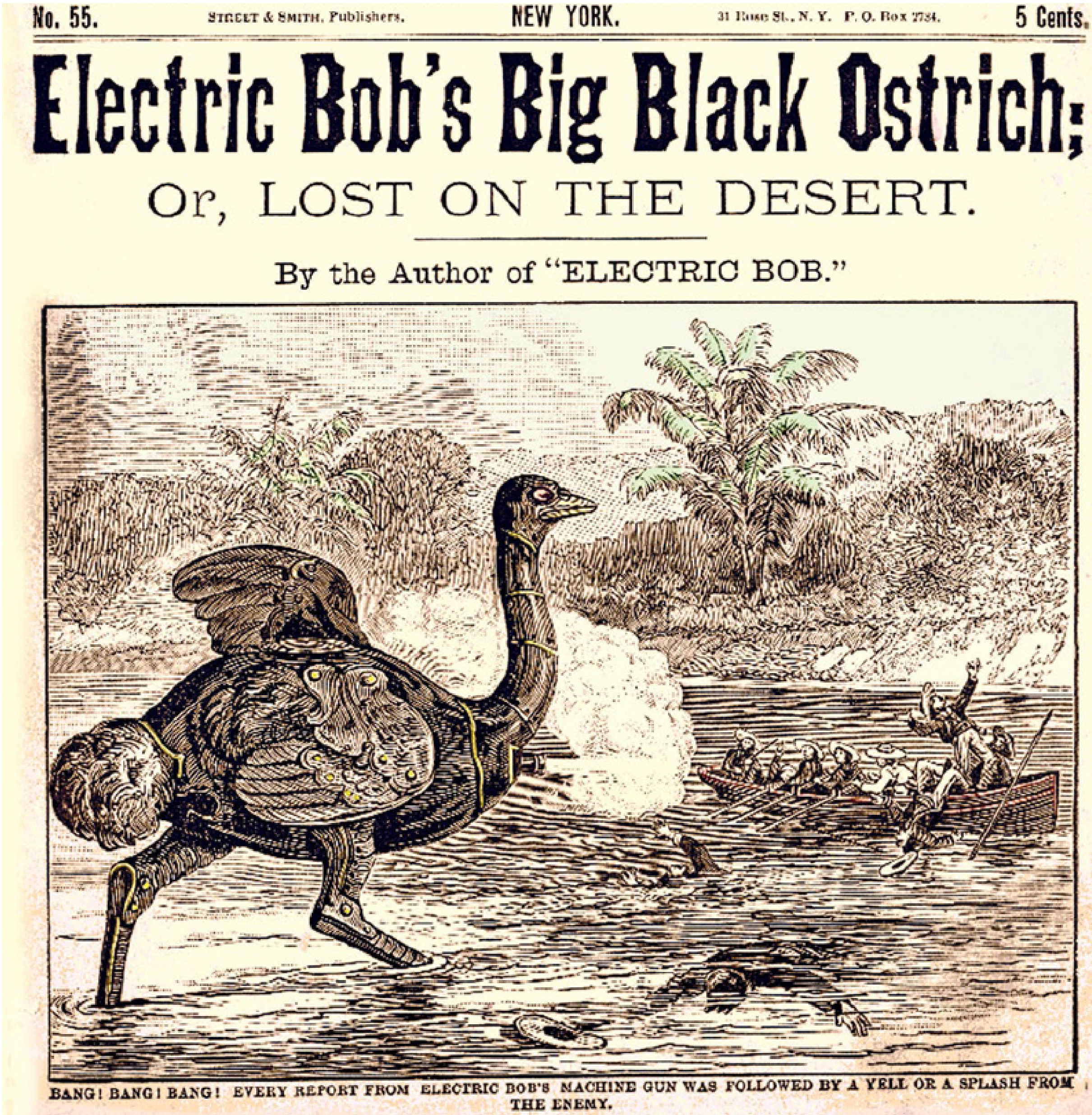
The Tower of Hanoi puzzle, and many variations, have been used in various robotics challenges, as the puzzle provides a useful standardized test to evaluate the robot's integration of high-level reasoning capabilities together with perception and manipulation.

Task planning and motion planning (involving one or more robot arms) play key roles in such challenges.

Simple algorithms exist for solutions involving three pegs, and the game is often used in computer programming classes to teach recursive algorithms. However, the optimal solution for the Tower of Hanoi problem (and variations) with large numbers of pegs is often unknown. For multi-armed robots, collision-free trajectories must also be computed.

SEE ALSO Tic-Tac-Toe (c. 1300 BCE), Mechanical Turk (1770), Connect Four (1988), Rubik's Cube Robots (2018)

Illustration for ***"Electric Bob's Big Black Ostrich,"*** a story by Robert T. Toombs in the *New York Five Cent Library*, 1893.



1893

ELECTRIC BOB'S BIG BLACK OSTRICH

The *Electric Bob* series of Stories by Robert T. Toombs, along with *The Steam Man of the Prairies* series, are notable for showing a growing fascination at the end of the nineteenth century in the United States with mechanical devices that resemble humans or animals. In *Electric Bob's Big Black Ostrich* (1893), Electric Bob is a ten-year-old engineering genius who lives near New York City and is a descendant of telegraph inventor Samuel Morse (1791–1872). Aside from his electronic ostrich, Bob is also an inventor of a giant mechanical white alligator and other mechanized animals that could be used for transport. Generally, the various robots are well stocked with supplies, have armor, and are capable of traversing difficult environments.

During the course of the story, Electric Bob reasons that a large electronic ostrich could carry him and his friends across a rocky desert in the American Southwest, while avoiding snakes. Bob carefully studies anatomy and physiology from live ostrich specimens and designs the perfect conveyance: "The ostrich towered thirty feet in the air to the top of his great head. The center of the body was twenty feet from the ground, the neck was about eight feet long. . . . The power is furnished by powerful storage batteries placed in the body just between the thighs of the bird, and

are capable of giving us a speed of twenty to forty miles an hour—depending on the nature of the ground we travel over."

The novel is notable for the engineering details it provides to readers. For example, it breaks down all the various materials and components of the mechanical bird, including its hollow steel legs and the bulletproof aluminum wings and tail. Bob continues: "Here are the water tank, storage placed for provision, ammunition, etc., and here is our machine gun . . . [which] consists of an enlarged revolved cylinder, holding twenty-five Winchester rifle cartridges, and a short, heavy barrel, and is fired by turning this crank."

Although such fictional treatments do not really address philosophical issues revolving around the creation of such mechanical life forms, Edisonades like this one provide a glimpse into the thinking of the time, along with all its prejudices, hopes, and aspirations.

SEE ALSO de Vaucanson's Duck Automaton (1738), *The Steam Man of the Prairies* (1868), Tik-Tok (1907)

Nikola Tesla's 1898 patent for his wirelessly controlled robotic boat, which contained a battery, motor-driven propeller, rudder, and lights. He believed that intelligent "teleautomatons" would be created in the future, causing a revolution in society.

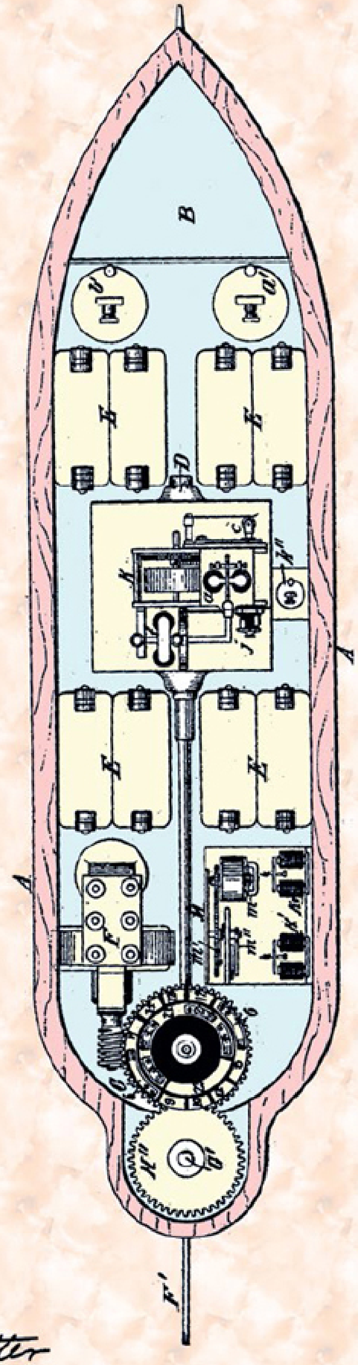
No. 613,809. Patented Nov. 8, 1898.

N. TESLA.

METHOD OF AND APPARATUS FOR CONTROLLING MECHANISM OF MOVING VESSELS OR VEHICLES.

(No Model.) 5 Sheets—Sheet 1.

Fig. 1



Witnesses:
Raphaël Netter
George Scherff.

Inventor
Nikola Tesla

1898

TESLA'S "BORROWED MIND"

In 1898, Serbian-American inventor Nikola Tesla (1856–1943) gave a remarkable demonstration of a radio-controlled boat, which he maneuvered in front of awestruck spectators, some of whom thought magic, telepathy, or a trained monkey might be involved. When a *New York Times* reporter learned of this world's-first radio-controlled vessel, he suggested that Tesla's invention could serve as a weapon of war, carrying dynamite. Tesla told the reporter to think beyond wireless torpedoes and realize that this represented the first of a race of automatons (the word *robot* was not yet used)—mechanical men that would perform the laborious work of the human race.

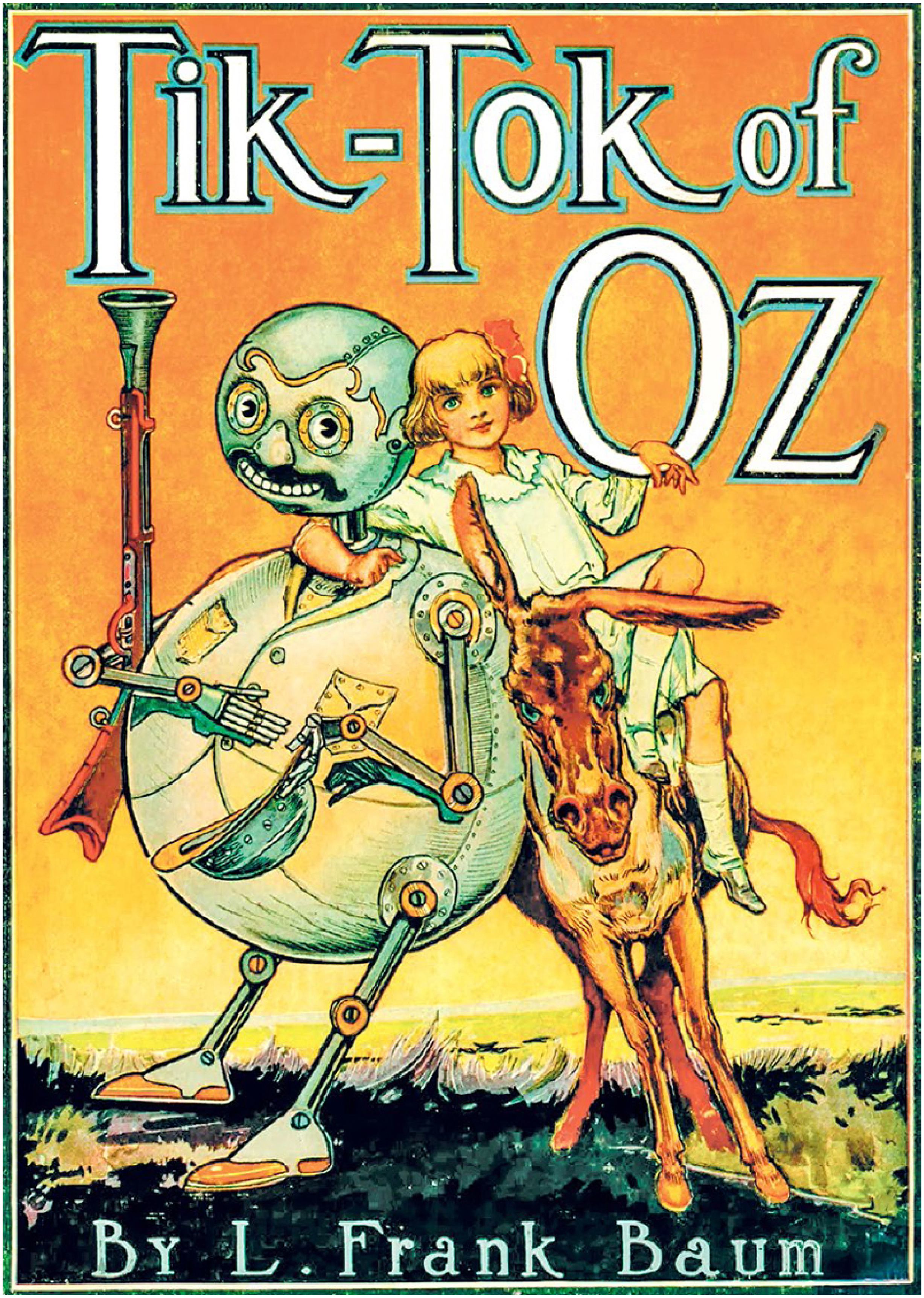
In his 1900 essay "The Problem of Increasing Human Energy," Tesla wrote about his aquatic automatons: "The knowledge, experience, judgment—the mind, so to speak—of the distant operator was embodied in that machine, which was thus enabled to move and to perform all its operations with reason and intelligence. . . . The automatons so far constructed had 'borrowed minds,' so to speak, as each merely formed part of the distant operator."

Tesla went even further to suggest that "an automaton may be contrived that will have its 'own mind' [and be] able, independent of any operator, left entirely to itself, to perform, in response to external influences affecting its sensitive organs, a great variety of acts and operations as if it had intelligence."

Tesla is often considered a great prophet of our modern technological age, where realistic androids (i.e., humanoid robots) like Nadine and Sophia baffle spectators with their ability to carry on conversations. In fact, Tesla believed that humans were mere automatons that responded to external stimuli and thought and acted accordingly, and among his grand ideas was the construction of "an automaton which would mechanically represent me, and which would respond, as I do myself, but, of course, in a much more primitive manner, to external influences. Such an automaton evidently had to have motive power, organs for locomotion, directive organs, and one or more sensitive organs so adapted as to be excited by external stimuli. . . . Whether the automaton be of flesh and bone, or of wood and steel, it mattered little, provided it could perform all the duties required of it like an intelligent being."

SEE ALSO Lethal Military Robots (1942), *The Human Use of Human Beings* (1950), Autonomous Vehicles (1984)

Cover of *Tik-Tok of Oz* by L. Frank Baum, 1914, illustrated by John R. Neill (1877–1943).


Tik-Tok of Oz
By L. Frank Baum

1907

TIK-TOK

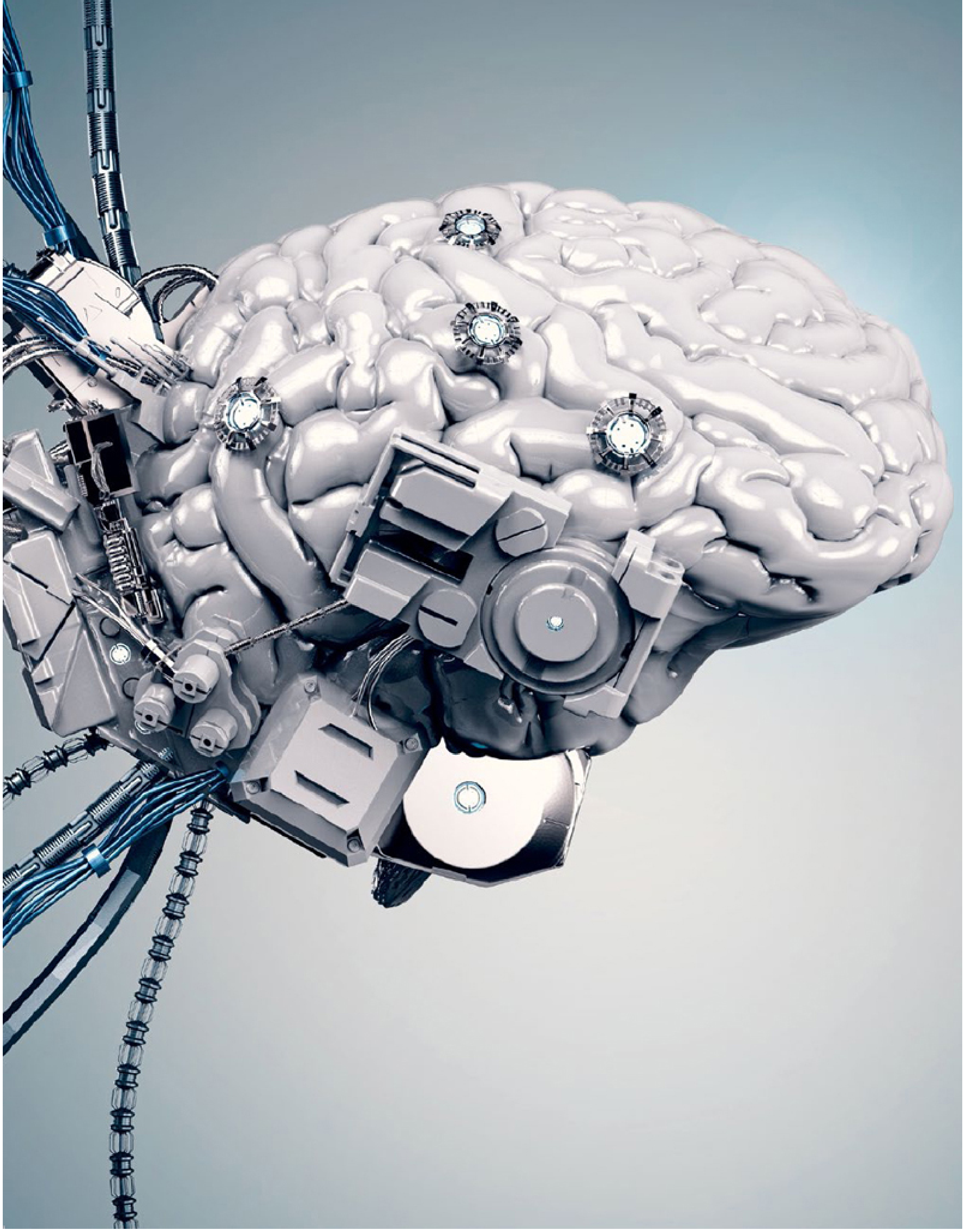
When we study artificial intelligence, “we are immediately confronted with the most fundamental inquiries into the nature of life, death, sex, work, and the mechanisms of the mind,” write authors Paul Abrahm and Stuart Kenter, “. . . a large order requiring exhausting pilgrimages into literature, philosophy, and an impressive array of scientific and technical realms.” One of the earliest thinking machines in literature to raise questions regarding the fine line between machines and people is Tik-Tok, an intelligent copper robot introduced in the 1907 novel *Ozma of Oz* by American author L. Frank Baum (1856–1919). To power the clockwork robot, someone must periodically wind up his three clockwork springs that separately energize his thoughts, actions, or speaking. For example, it is possible to energize his thoughts but not his actions or speech, creating an isolated “AI in a Box.” Or his speech may be energized, but not his thoughts, allowing him to create crude vocal output but not having adequate natural language-processing facilities. Even when fully activated, his language-processing modules are not entirely natural, as evidenced by his monotonous-sounding voice and strict literal interpretation of many questions and commands. According to Baum, Tik-Tok “does everything but live” and feels no emotions. When punished with whips, Tik-Tok is not harmed, as the whipping merely kept his copper body “well polished.”

Tik-Tok realizes his place in the universe. For example, when thanked for his kindness, he replies “I am only a machine. I cannot be kind any more than I can be sorry or glad.” Even though the novel appears to be targeted to a youthful readership, Baum makes us wonder about the future of AI. Is emotion the main differentiator between humans and ma-chines? To what degree do literature and movies shape AI designs and the limitations we may place on thinking machines?

“Cyborgs, robots and other mechanical beings are key figures for understanding the technophilic and technophobic dreams of a century,” writes Professor Alex Goody, “embodying fears about technological encroachment, suggesting to some the chance for technological transcendence, and challenging the idea of the individual, differentiated, human subject.”

SEE ALSO Lancelot’s Copper Knights (c. 1220), *The Steam Man of the Prairies* (1868), Asimov’s Three Laws of Robotics (1942), Natural Language Processing (1954), Leakproof “AI Box” (1993)

Alan Turing wrote that humans would not be irreverently usurping God's power of creating souls when we someday create advanced thinking machines—any more than we do when we produce children.



1907

SEARCHES FOR THE SOUL

In his 1950 paper "Computing Machinery and Intelligence," computer scientist Alan Turing (1912–1954) wrote that when we attempt to construct AI machines, we are not "irreverently usurping" God's "power of creating souls, any more than we are in the procreation of children; rather we are, in either case, instruments of His will, providing mansions for the souls that He creates." Some futurists believe that, as we learn more about the structure of the brain, conscious AIs might be created by simulating a mind or by uploading aspects of our minds to a computer. These speculations assume a materialist view in which the mind arises from brain activity. On the other hand, in the mid-seventeenth century French philosopher René Descartes (1596–1650) supposed that the mind, or "soul," exists separately from the brain. In his view, this "soul" is connected to the brain via an organ like the pineal body, which acts as a portal between brain and mind.

The various views concerning the separation of soul and matter represent a philosophy of mind-body *dualism*. In 1907, American physician Duncan MacDougall (1866–1920) placed dying tuberculosis patients on a scale in an attempt to demonstrate this concept. He reasoned that at the moment of death, the scale should indicate a drop in weight as the soul disembarked. As a result of his experiments, MacDougall measured the soul to be 21 grams (0.7

ounces). Alas, MacDougall and other researchers were never able to duplicate this finding.

A more materialist view of the mind and body may be supported by experiments that suggest our thoughts, memory, and personality can be altered by damage to regions of the brain, and brain-imaging studies can map both feelings and thoughts. As just one curious example, injury to the brain's right frontal lobe can lead to a sudden, passionate interest in fine restaurants and gourmet foods—a condition called *gourmand syndrome*. Of course, the dualist Descartes might have argued that damage to the brain alters behavior because it is through the brain that the mind operates. If we excise the car's steering wheel, for example, the car behaves differently, but this does not mean that there is no separate driver.

SEE ALSO The Consciousness Mill (1714), Transhumanism (1957), Living in a Simulation (1967), Spielberg's *A.I. Artificial Intelligence* (2001)

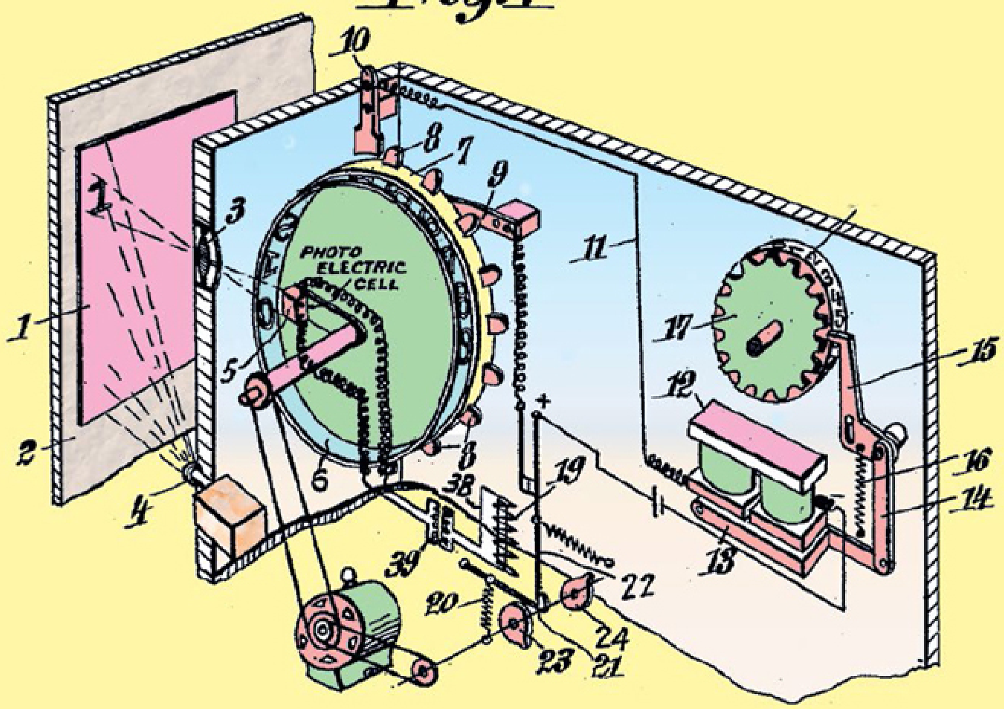
Illustrated here is a reading machine by Austrian engineer Gustav Tauschek (1899–1945), from US Patent 2,026,329. The comparison device (disk 6) had cutouts for letters. When images of a character and a letter-shaped hole coincided, the appropriate letter was printed.

Dec. 31, 1935. | G. TAUSCHEK | 2,026,329

READING MACHINE

Filed May 27, 1929

Fig. 1



1913

OPTICAL CHARACTER RECOGNITION (OCR)

President John Adams is often credited with saying “The only thing most people do better than anyone else is read their own handwriting.” Indeed, the quest for automated systems that could read printed letters of the alphabet has had a long history. *Optical character recognition*, or OCR, has involved several fields of research and development: computer vision, artificial intelligence, pattern recognition, and more. OCR refers to the conversion of images of text in numerous varieties (e.g., handwritten, printed, typed) into machine-encoded text. For example, the text on a mailing envelope, license plate, book page, street sign, or passport might be scanned and recognized by a machine. Sometimes OCR is used to convert text to speech for blind people.

One early inventor in the area of OCR is Russian-born scientist Emanuel Goldberg (1881–1970), who in 1931 patented a document-search device that used photoelectric cells and pattern recognition to search for information on microfilmed documents. Even earlier, around 1913, Irish physicist Edmund Fournier d’Albe (1868–1933) invented the *optophone*, which assisted blind readers by using photosensors to scan text and generate tones that corresponded to letters. In 1974, American inventor Ray Kurzweil (b. 1948) created a reading machine for the blind that could scan text in numerous different fonts and produce speech output.

OCR often requires a number of interesting processing steps, such as mathematically tilting the text as needed, removing noise, and smoothing edges. In order to recognize characters, the system may make comparisons to stored characters and/or consider particular graphical features (e.g., loops and lines).

A closely related field includes *handwriting recognition* (HWR), which may also involve the monitoring and analysis of the movements of a pen in recognizing the word being written. HWR usually employs OCR, but may also involve determining the most plausible words in a given context to increase accuracy. Artificial neural networks can also be used to facilitate OCR and HWR.

SEE ALSO Artificial Neural Networks (1943), Speech Recognition (1952), Machine Learning (1959)

Poster for a production of *R.U.R.,* directed by the puppeteer Remo Bufano in New York, in 1939. The performance was funded by the Federal Theatre Project (1935–1939), a New Deal program to fund art in the United States during the Great Depression.


DIRECTOR

1920

ROSSUM'S UNIVERSAL ROBOTS

Written by the Czech art critic and playwright Karel Čapek (1890–1938), the 1920 play *R.U.R.* (*Rossum's Universal Robots*) introduced the word *robot* to the English language.

In *R.U.R,* the robots are made of flesh and blood but assembled in vats. They serve humankind as factory workers—essentially as inexpensive appliances—which allows humans a tremendous amount of leisure time. However, debates take place about their rights and humanity. Helena, one of the main characters, desires to free the robots. Alas, the robots, which are being used worldwide, eventually destroy the human race; but because the robots do not possess the secret formula to reproduce themselves, they too will eventually die out. At the end of the play, two special robots fall in love, representing a kind of Adam and Eve for the future of our planet.

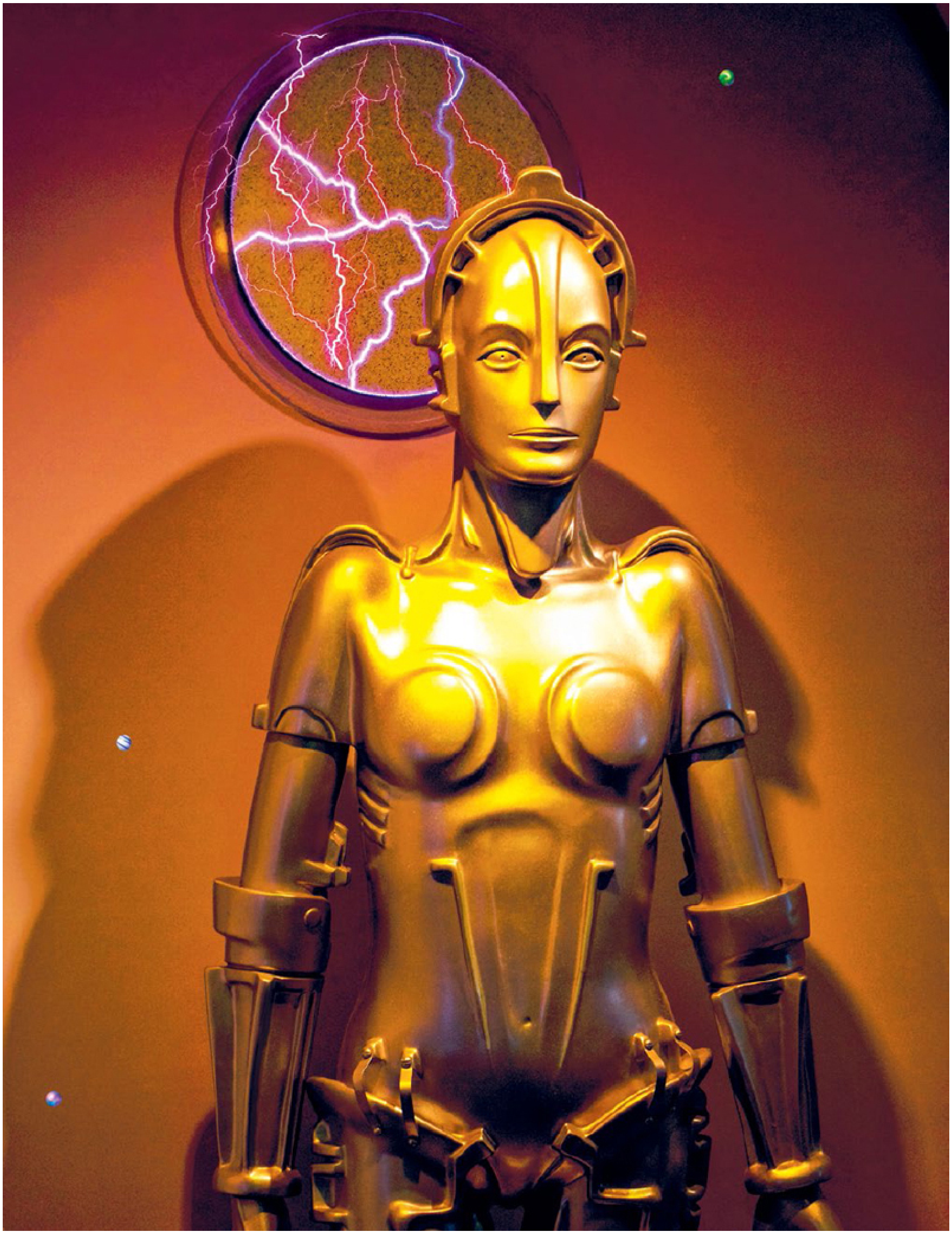
The word *robot* comes from the Czech *robota*, which refers to forced labor. The play is a milestone in that it forced people to consider the implications of the continued evolution of AI technology—not only in the area of jobs and possible dehumanization effects on society, but also regarding the general safety of the human race. Where do we draw the line between humans and thinking machines? When do such machines become so advanced that they deserve rights or become a threat? According to author Rebecca Stefoff,

R.U.R. suggests that the "indefinable spark of humanity dwells in how beings feel and act, not in how they are created."

"Philosophically rich and controversial, *R.U.R.* was unanimously acknowledged as a masterpiece from its first appearance, and has become a classic of technologically dystopian literature," writes Luciano Floridi (b. 1964), Professor of Philosophy and Ethics of Information at Oxford University. The ideas in the play were so powerful that, by 1923, the play had been translated into more than thirty languages. In 1922, the American premiere of the play took place in New York City, where it had more than a hundred performances.

SEE ALSO *Metropolis* (1927), *The Human Use of Human Beings* (1950), Intelligence Explosion (1965), Leakproof "AI Box" (1993)

Maria from the film *Metropolis,* on display at the Robot Hall of Fame at the Carnegie Science Center in Pittsburgh, Pennsylvania.



1927

METROPOLIS

In the 1927 silent film *Metropolis*—directed by Fritz Lang (1890–1976) and scripted by Thea von Harbou (1888–1954)—the inventor C. A. Rotwang explains that his robots never tire or make a mistake and that these workers of the future will not be distinguishable from human beings. Set in a futuristic city, humankind is divided into a leisure class that rules the city and the underclass that runs the vast machinery, toiling below ground.

Maria, the heroine, is a young woman who cares about the workers and the difficulty of their lives. As the visually impressive film progresses, Rotwang creates a robot that resembles Maria in an attempt to ruin her reputation among the workers and discourage any rebellion. The simulated Maria actually urges the workers to revolt, but she is later captured and burned at the stake. As she burns, the human-like exterior melts away, revealing the metallic-looking robot within.

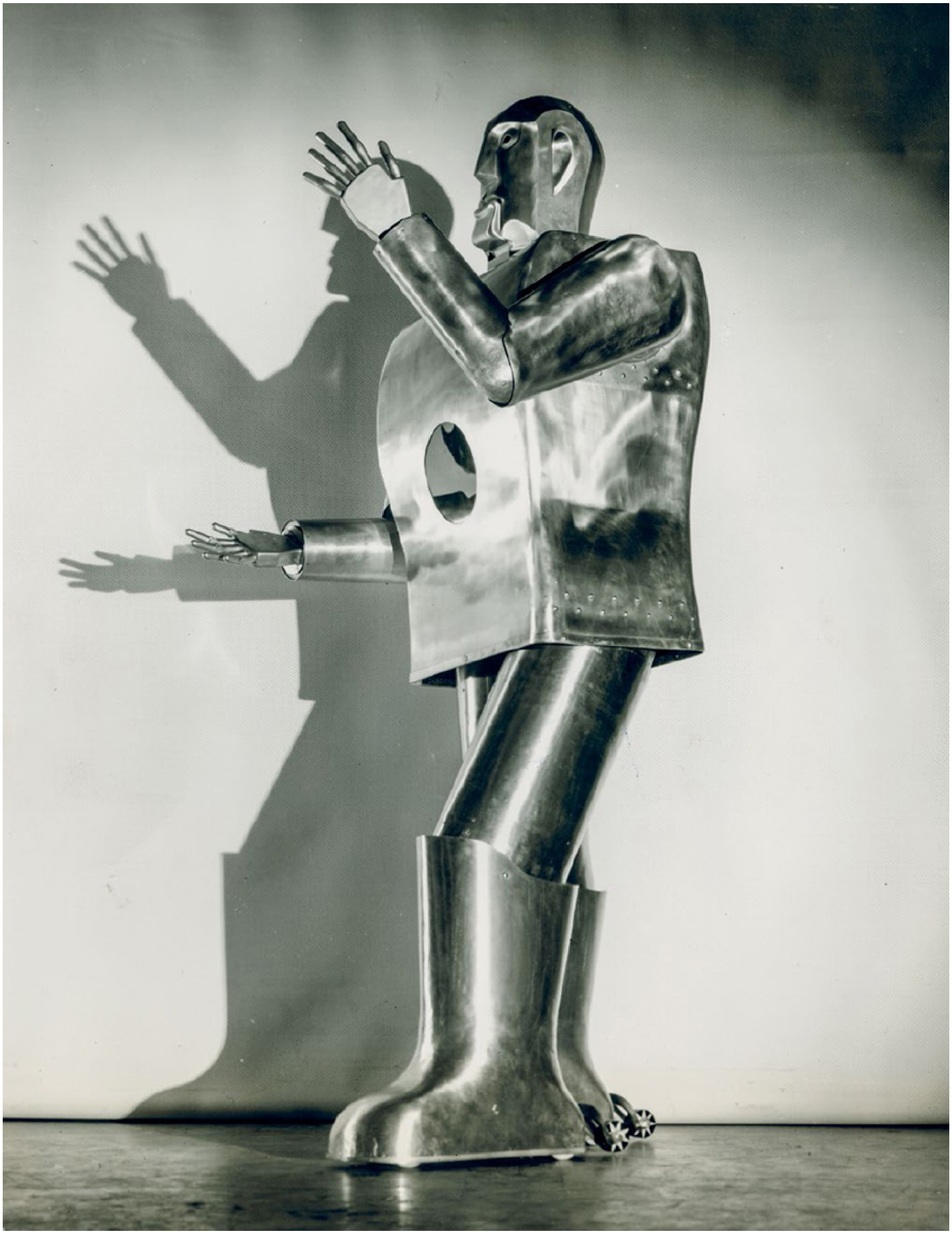
The film's motto is "The mediator between the brain and muscle must be the heart," and as such it probes the essential differences between humans and AI. Considering the ramifications of *Metropolis*, the futurist Thomas Lombardo writes: "The robot, in fact, in science fiction emerged as a symbolic synthesis of humanity and the machine—equally the human becoming machine-like, being assimilated by its technological creation, as well as the machine becoming human, embodying our worst qualities and characteristics.

. . . It personifies our fear of science and technology, as well as our fear of what we may become."

The themes of *Metropolis* echo in more modern science-fiction films like the cult classic *Blade Runner* (1982) with its synthetic humans. Of course, overreliance on technology, and the future of labor in an AI age, are hot topics of discussion today. *Metropolis* will continue to be relevant in the future, as AI entities are increasingly mistaken for humans. When AIs are developed with the ability to impersonate people we trust or respect, or to become "simulated beings" with whom we fall in love, the consequences will be far-reaching.

SEE ALSO Rossum's Universal Robots (1920), Turing Test (1950), Ethics of AI (1976), *Blade Runner* (1982)

Elektro, the "Moto-Man," from the Westinghouse Exhibit at the 1939–1940 New York World's Fair.



1939

ELEKTRO THE MOTO-MAN

Elektro deserves a place in this book because he has been hailed as one of the world's first "celebrity robots" and "America's oldest surviving robot."

He was built by the Westinghouse Electric Corporation and exhibited at the 1939 New York World's Fair, where he was an instant hit. This 7-foot-tall (2.1-m) humanoid could move in response to voice commands, could speak hundreds of words, and even smoked cigarettes. His photoelectric eyes could distinguish red and green light. In 1940, he was accompanied by Sparko, a robot dog that could bark and move. Countless visitors to the World's Fair stood in line to see his twenty-minute performances.

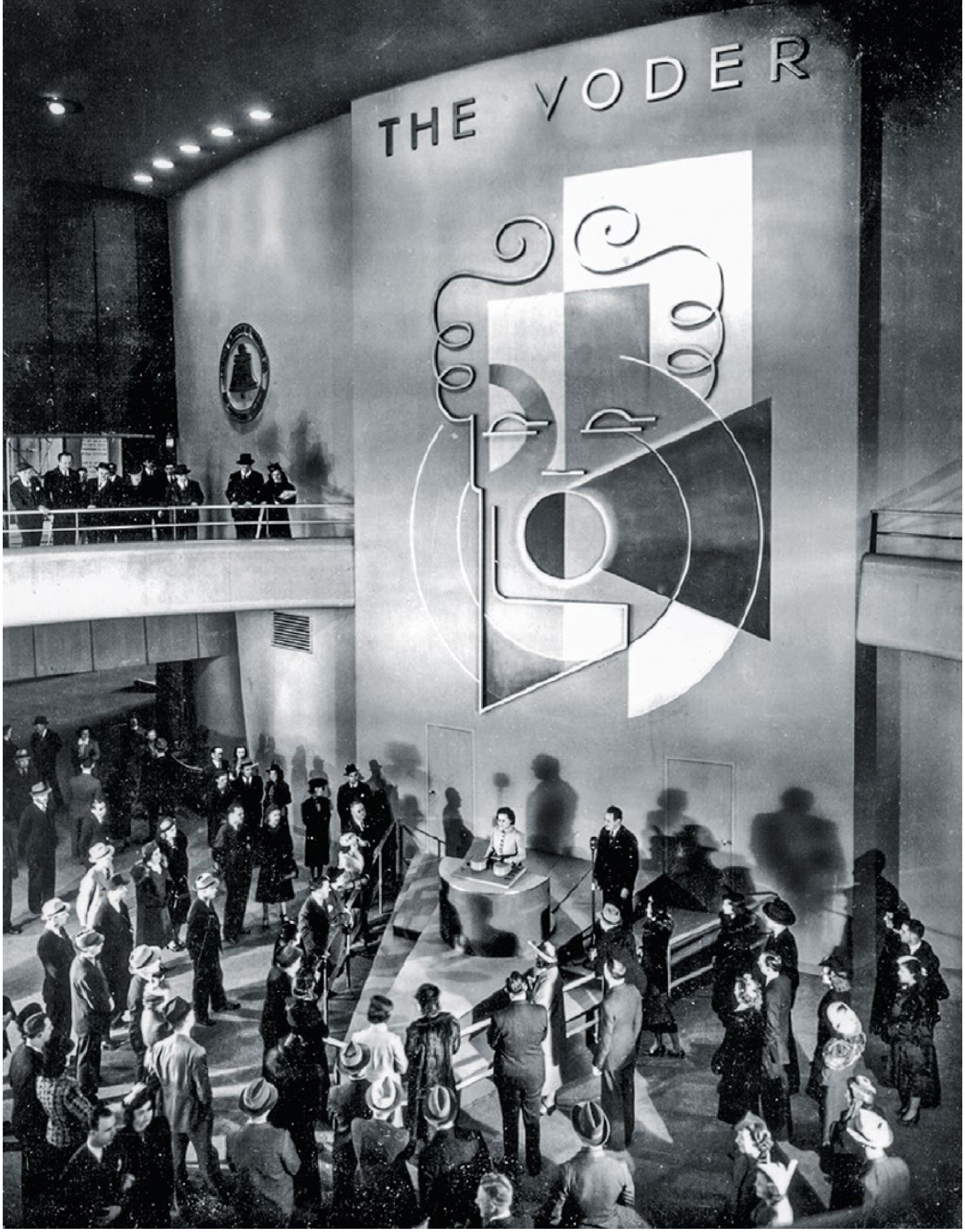
Realizing that many people would mistakenly consider his capabilities to be a result of a hidden man inside a robot costume, his designers deliberately cut a hole in his body to show that this was not the case. In actuality, he was created with camshafts, gears, and motors that moved his head, mouth, and arms. Engineer Joseph Barnett invented him, making use of a series of 78-rpm record players connected to relay switches to generate his 700-word vocabulary. The robot could say things like "My brain is bigger than yours," and he responded to speech based on the number of words or syllables he heard. For example, three words (no matter what they were) activated relays to stop Elektro's motion. Alas, he and Sparko

could not wander very far, as they were actually controlled by nearby operators via electrical cables connected near the feet.

Elektro has inspired a number of children to pursue careers in engineering over the years. He also appeared in the 1960 comedy film *Sex Kittens Go to College* and then was promptly dismantled, with his head given as a retirement gift to a Westinghouse employee. In 2004, his various pieces were rediscovered, and he was finally reassembled.

SEE ALSO da Vinci's Robot Knight (c. 1495), *The Steam Man of the Prairies* (1868), Shakey the Robot (1966), ASIMO and Friends (2000)

The VODER, which simulated effects of the human vocal tract, fascinated crowds at the 1939 New York World's Fair. The VODER employed a console from which an operator could create speech.


THE VODER

1939

SPEECH SYNTHESIS

Perhaps many of you have heard the synthesized voice of astrophysicist Stephen Hawking (1942–2018), who for many years used a speech synthesizer with a robotic voice to speak on his behalf after his motor neuron disease led to his inability to speak naturally. In fact, the ability of computer systems to convert text to speech has served many useful purposes, including reading text out loud for the visually impaired, young children, or people with various reading difficulties. Synthetic speech can also allow computer systems to create the impression of human intelligence and interaction when interfaced with digital personal assistants of all kinds. New approaches using neural nets can now simulate the specific natural speech of selected people, and thus we live in a world where it will be increasingly difficult to know when the voice of someone we trust (e.g., a business partner, a parent, a child) is being generated by the actual person. What happens when someone can "steal" anyone else's voice and have them speak anything desired?

Speech synthesis systems have been implemented by various means. For example, engineers can store digitized speech units and concatenate them (i.e., string them together) during playback, or they can utilize distinctive frequency components of the acoustic signal (i.e., formants) in a method called *formant synthesis*. With *articulatory synthesis*, models of the human vocal tract are

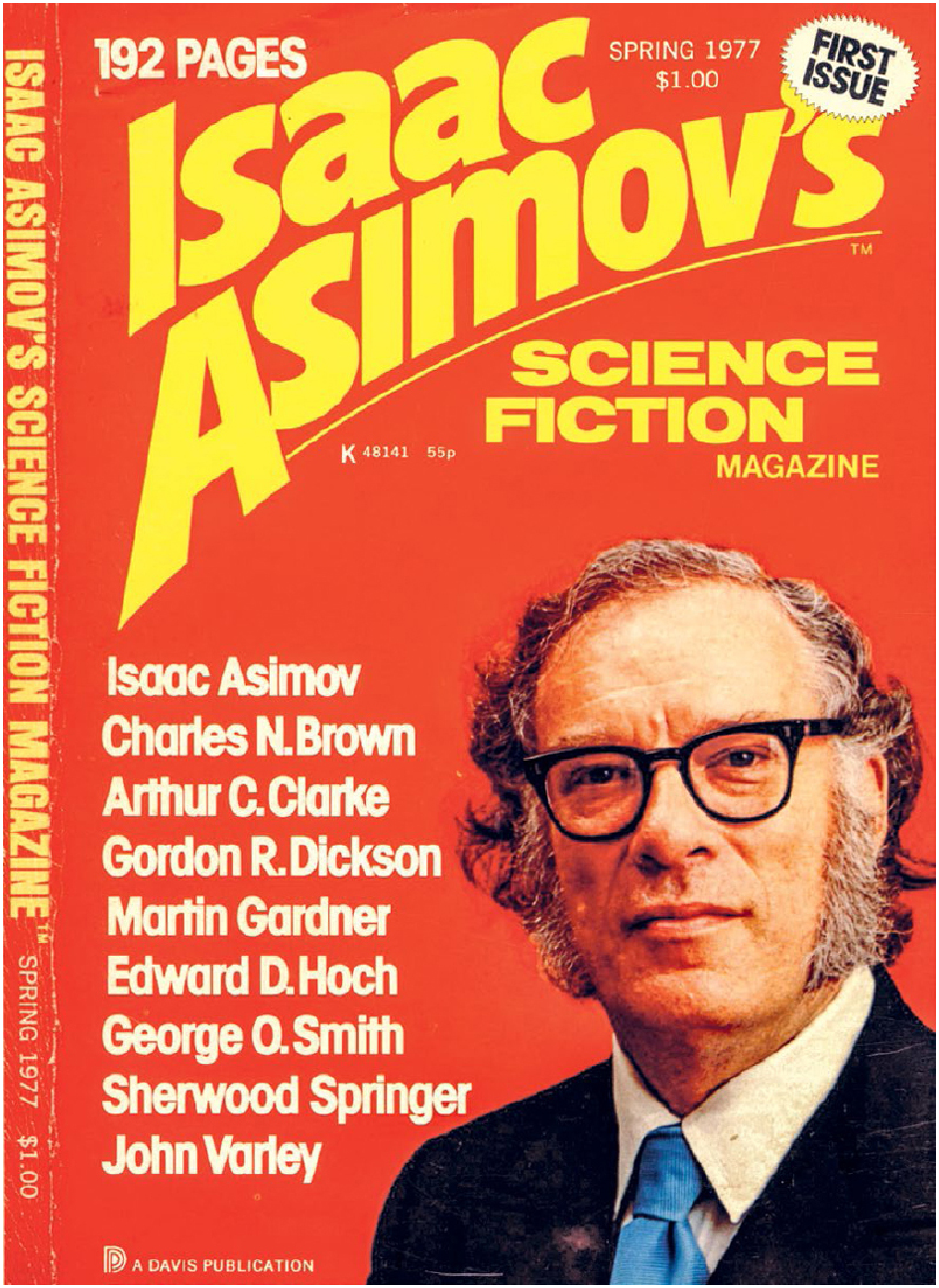
simulated. Of course, simple systems like talking clocks, cars, toys, and calculators, need only store a few pre-recorded words for playback.

Many challenges exist for converting text to natural and understandable speech. For example, consider the various pronunciations in the English language of words like *tear*, *bass*, *read*, *project*, *desert*, and so on, which change depending on the context in which they are spoken.

One of the early landmarks in the history of synthetic speech involves the work of engineer Homer Dudley (1896–1980), who invented the *vocoder* (i.e., “voice encoder”), which could electronically produce speech using various electronic filters, and the *VODER* (i.e., “voice operation demonstrator”), which used a console from which an operator could create speech. The latter, which simulated effects of the human vocal tract, was exhibited at the 1939 New York World’s Fair.

SEE ALSO Speech Recognition (1952), Natural Language Processing (1954), Ethics of AI (1976)

Author Isaac Asimov, famous for his “Three Laws of Robotics,” is shown here in 1977 on the cover of his science-fiction magazine, in which he further develops the idea of artificial intelligence in his story “Think!”


ISAAC ASIMOV'S SCIENCE FICTION MAGAZINE™ SPRING 1977 $1.00
192 PAGES
SPRING 1977
$1.00
FIRST ISSUE
Isaac Asimov's™
SCIENCE FICTION
MAGAZINE
K 48141 55p
Isaac Asimov
Charles N. Brown
Arthur C. Clarke
Gordon R. Dickson
Martin Gardner
Edward D. Hoch
George O. Smith
Sherwood Springer
John Varley
A DAVIS PUBLICATION

1942

ASIMOV'S THREE LAWS OF ROBOTICS

As AI and robotics advance in the coming decades, what constraints or codified laws should be developed to ensure that such entities do not take actions that harm humans? In 1942, author and edu-cator Isaac Asimov (1920–1992) introduced his famous "Three Laws of Robotics" in a short story called "Runaround," which features a smart robot's interactions with people. The Three Laws are (1) a robot may not injure a human being or, through inaction, allow a human being to come to harm; (2) a robot must obey the orders given it by human beings except where such orders would conflict with the First Law; (3) a robot must protect its own existence as long as such protection does not conflict with the First or Second Laws. Asimov went on to write many stories that illustrated how these simple laws could have unintended consequences.

Later, he provided an additional law: "A robot may not harm humanity, or, by inaction, allow humanity to come to harm." These laws have been influential, not only for science-fiction writers, but also for AI experts. AI researcher Marvin Minsky (1927–2016) noted that after encountering Asimov's laws, he "never stopped thinking about how minds might work. Surely we'd someday build robots that think. But how would they think and about what? Surely logic might work for some purposes, but not for others. And how to build robots

with common sense, intuition, consciousness and emotion? How, for that matter, do brains do those things? ”

The laws are notable and useful for the countless questions they raise. What other laws might we add to the Asimov set? Should robots never pretend they are human? Should robots “know” that they are robots? Should they always be able to explain why they acted as they have? What if a terrorist used multiple robots to harm people, without each robot knowing the entire plan, and thus not violating the First Law? We may also consider how these laws may have an impact on robot army medics who must perform triage when they cannot tend to multiple injured soldiers, or autonomous vehicles that must determine whether to crash into playing children or drive off a cliff and kill a passenger. Finally, could a robot really decide what it means to “harm humanity,” given that its interactions could have repercussions for years into the future?

SEE ALSO Lethal Military Robots (1942), Ethics of AI (1976), *Blade Runner* (1982), Autonomous Vehicles (1984)

Artist's conception of a lethal autonomous drone, attacking enemy tanks after visual recognition and AI confirmation.

1942

LETHAL MILITARY ROBOTS

Robots used in warfare include many examples in the 1900s, such as the tank-like Goliath used by the Germans beginning in 1942, during World War II, on all fronts where the Wehrmacht fought. The unmanned Goliaths were remote-controlled via a connecting cable and carried high explosives so that the machines could blow up along with their intended targets.

Today, drones (unmanned flying vehicles) can be armed with missiles to serve as an effective weapons system, but they usually require remote human input and authorization before being “allowed” to destroy a target. Of historical note, an MQ-1 Predator drone launched the first-ever deadly airstrikes from a drone in 2001 in Afghanistan. Debates continue regarding the possible future use of lethal *autonomous* weapons that could actually select and attack military targets without human intervention. Automatic defensive systems do exist today, including machines that can autonomously identify and attack incoming missiles.

Possible advantages of military robots are many: they never tire or show fear; they can rapidly perform maneuvers that would injure human pilots; and they could potentially save lives of soldiers and reduce collateral damage and civilian casualties. In principle, machines might be instructed to follow various rules, such as not to fire when unsure if a target is a civilian or a combatant, or whether lethal force is permitted. Perhaps damage potential to civilians might

be programmed and restricted to be proportional to the size of a military target. Facial-recognition software could be used to enhance accuracy, and military robots could work side by side with soldiers, enhancing their abilities much like software or robotics may be used today to assist with surgeries. But how much independence should such fighting machines be given? Who is at fault if a robot accidentally attacks a school?

In 2015, a large group of AI experts signed a letter warning of the dangers of military use of offensive autonomous weapons beyond human control, which could lead to a global AI arms race. The letter was presented at the International Joint Conference on Artificial Intelligence and signed by such intellectuals as Stephen Hawking, Elon Musk, Steve Wozniak, and Noam Chomsky.

SEE ALSO Tesla's "Borrowed Mind" (1898), Asimov's Three Laws of Robotics (1942), *Colossus: The Forbin Project* (1970), Ethics of AI (1976), Autonomous Robotic Surgery (2016), Adversarial Patches (2018)

Artificial neural networks are loosely inspired by biological neural networks, such as interconnected neurons in the brain that transmit signals to one another.

1943

ARTIFICIAL NEURAL NETWORKS

Artificial neural networks (ANNs) are often schematically represented like layer cakes, made from icing and stacked sheets of cake. For our purposes, the layers contain neurons in the form of simple computational units that become "excited" and propagate that excitement to other connected neurons. A weight, or strength, factor determines how much excitement should get passed along. Through a training period, as various weights and thresholds are adjusted from an initially random state, these systems can learn to perform tasks, such as recognizing an image of an elephant—by first analyzing numerous images that have been labeled as "elephant" and "not elephant." Various embellishments to the basic neural net function include "backpropagation," which can pass information in a reverse direction. Neural networks are now used in a wide variety of research and practical applications, including game playing, vehicle control, drug design, cancer detection in medical images, and language translation.

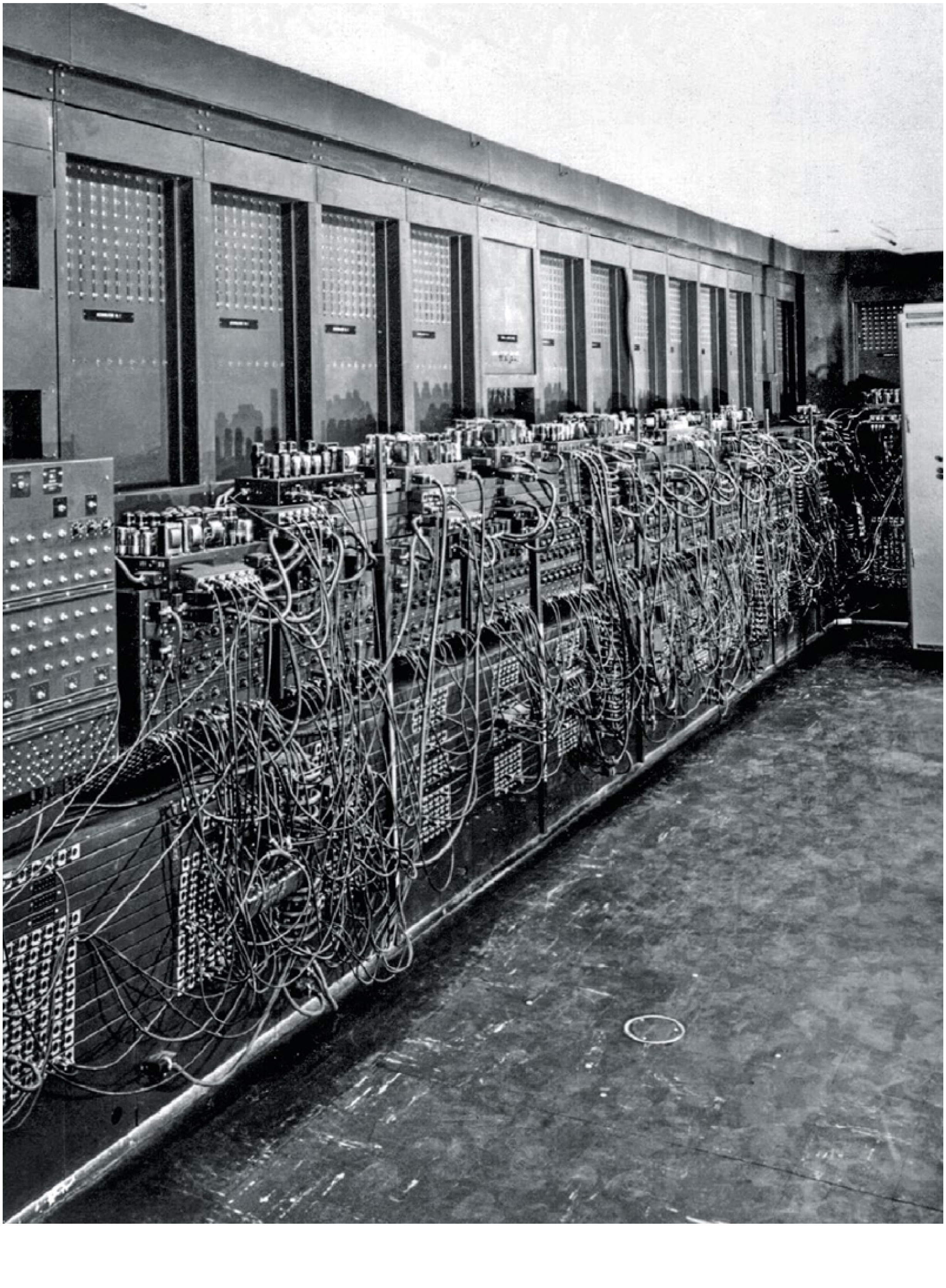
Some of the basic computational models in neural nets were discussed by neurophysiologist Warren McCulloch (1898–1969) and logician Walter Pitts (1923–1969) in their 1943 paper "A Logical Calculus of the Ideas Immanent in Nervous Activity," published in the *Bulletin of Mathematical Biophysics*. In 1957, Frank Rosenblatt (1928–1971) created the perceptron algorithm for pattern recognition, which was subsequently implemented in computer

hardware. In the twenty-first century, the utility of ANNs dramatically improved through the use of distributed computing (e.g., on different networked computers) and GPU (graphical processing unit) hardware.

ANNs are inspired by biological neural networks and are one means of implementing machine learning, in which computers exhibit the ability to learn without being explicitly programmed for a task. The concept of "deep learning" in computer science includes ANNs with several layers, which allow rich intermediate representations to be built. One challenge associated with ANNs is that input can sometimes be intentionally manipulated so that the ANNs are deceived into giving obviously incorrect answers. Also, it is difficult to understand how and why an ANN is providing a particular answer. Nevertheless, reflecting upon the recent useful applications of neural nets, Google AI expert Jeff Dean (b. 1968) said "The portion of evolution in which animals developed eyes was a big development. Now computers have eyes."

SEE ALSO Reinforcement Learning (1951), Perceptron (1957), Machine Learning (1959), Deep Learning (1965), Computer Art and DeepDream (2015)

ENIAC was among the first electronic, reprogrammable, digital computers that could be used to solve a large range of computing problems. The machine contained more than 17,000 vacuum tubes.



1946

ENIAC

The 1946 newspaper headlines were ecstatic about ENIAC and the future of thinking machines. “Mechanical Brain Enlarges Man’s Horizons,” wrote *The Philadelphia Inquirer*, and *The Cleveland Plain Dealer* proclaimed “Calculator Puts Humans to Shame”—and also said it heralded “a new epoch in the scope of human thought.” All this media attention about progress in electrically embodied “thinking,” along with the technological advancements, naturally led the world to consider further possibilities for AI.

ENIAC, short for Electronic Numerical Integrator and Computer, was built at the University of Pennsylvania by American scientists John Mauchly (1907–1980) and J. Presper Eckert (1919–1995). This device was among the first electronic, reprogrammable, digital computers that could be used to solve a large range of computing problems. The original purpose of ENIAC was to calculate artillery firing tables for the US Army. However, its first important application involved the design of the hydrogen bomb.

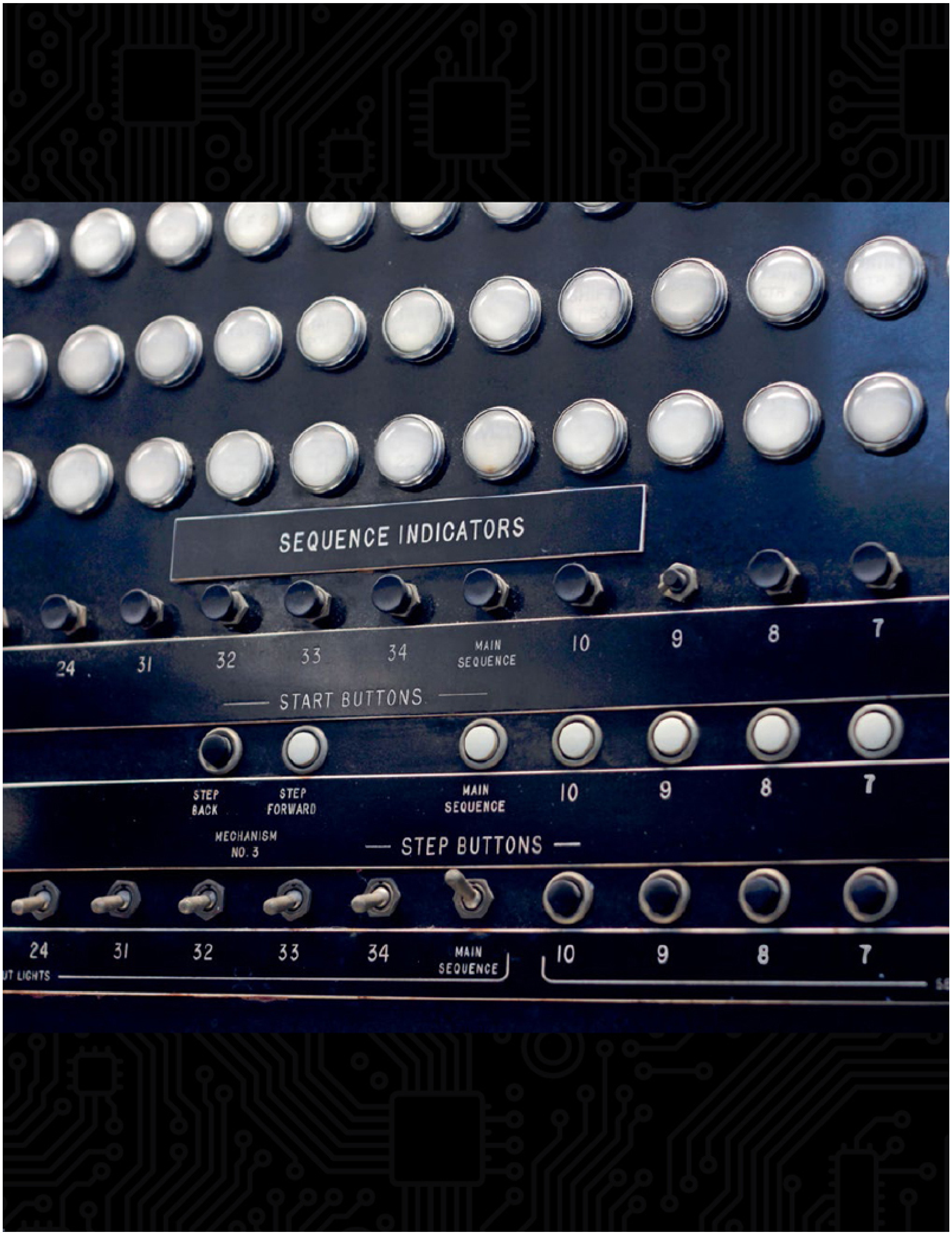
ENIAC was unveiled in 1946, having cost nearly $500,000, and it was in nearly continuous use until it was turned off on October 2, 1955. The machine contained more than 17,000 vacuum tubes and around five million hand-soldered joints. An IBM card reader and card-punch machine were used for input and output. In 1995, a team of engineering students led by Professor Jan Van der Spiegel created a “replica” of the 30-ton ENIAC on a single integrated circuit!

Other important electrical computing machines of the 1930s and 1940s include the American Atanasoff–Berry Computer (demonstrated in October 1939), the German Z3 (demonstrated in May 1941), and the British Colossus computer (demonstrated in December 1943). In contrast to ENIAC, however, these machines were either not fully electronic or not general-purpose.

The authors of the ENIAC patent (No. 3,120,606, filed in 1947) write: “With the advent of everyday use of elaborate calculations, speed has become paramount to such a high degree that there is no machine on the market today capable of satisfying the full demand of modern computational methods. . . . The present invention is intended to reduce to seconds such lengthy computations. . . .”

SEE ALSO Abacus (c. 190 BCE), Babbage’s Mechanical Computer (1822), *Giant Brains, or Machines That Think* (1949)

A “giant brain.” Shown here are the sequence indicators and switches of the IBM Automatic Sequence Controlled Calculator (ASCC), or Harvard Mark I Computer, in a science building at Harvard University.


SEQUENCE INDICATORS
24
31
32
33
34
MAIN SEQUENCE
10
9
8
7
START BUTTONS
STEP BACK
STEP FORWARD
MAIN SEQUENCE
10
9
8
7
MECHANISM NO. 3
STEP BUTTONS
24
31
32
33
34
MAIN SEQUENCE
10
9
8
7

1949

GIANT BRAINS, OR MACHINES THAT THINK

In 1949, the American computer scientist Edmund Berkeley (1909–1988) published what may be the first popular book on computers for a general audience. It was titled *Giant Brains, or Machines That Think*, and it was notable in that it raised questions about the appropriateness of using the term *brain* and the word *thinking* applied to computers. Many of these questions echo to this day. In the book, he writes: “Recently there has been a good deal of news about strange giant machines that can handle information with vast speed and skill. They calculate and they reason. Some of them are cleverer than others—able to do more kinds of problems . . . they can solve problems that a man’s life is far too short to permit him to do. . . . These machines are similar to what a brain would be if it were made of hardware and wire instead of flesh and nerves. It is therefore natural to call these machines mechanical brains.”

It is fascinating to realize that when the book was written, electronic computers were virtually unknown to the public. There were only a small number of these “giant brains” in existence, and Berkeley discusses several of them in the book, including the Differential Analyzer Number 2 at MIT, the Mark I at Harvard University (also known as the IBM Automatic Sequence Controlled Calculator), ENIAC at the Moore School, the General-Purpose Relay Calculator at Bell Laboratories, and the Kalin-Burkhart Logical-Truth

Cal-culator built by Harvard students, to name a few. In Berkeley's endnotes, added to the book in 1961, he noted that even "intuitive thinking" might be achieved by machines someday: "Perhaps intuitive thinking is a very rapid running over of possible alternatives in one's mind, coupled with very rapid appraising of them, so that one arrives at a conclusion with hardly any consciousness of how that conclusion was obtained. If so, then of course it is possible to program computers in such a way that they display what we call intuitive thinking, except that the method of obtaining the conclusions is known."

SEE ALSO Hobbes's *Leviathan* (1651), The Consciousness Mill (1714), ENIAC (1946), *Colossus: The Forbin Project* (1970), "Call Them Artificial Aliens" (2015)

The Turing test probes a machine's capability for exhibiting intelligent behavior indistinguishable from that of a person.

1950

TURING TEST

French philosopher Denis Diderot (1713–1784) once remarked: "If they find a parrot who could answer to everything, I would claim it to be an intelligent being without hesitation." This leads to the question: Can appropriately programmed computers be considered to be intelligent entities that "think"? In 1950, English computer scientist Alan Turing (1912–1954) attempted to answer this with his famous paper "Computing Machinery and Intelligence," published in the journal *Mind*. He suggested that if a computer behaves in the same way as a human, we might as well call it intelligent, and then he proposed a special test to evaluate any given computer's intelligence: Imagine that a computer and human respond, by text, to typed questions of human judges who cannot actually see who or what is responding. If the judges cannot distinguish the computer from the person after studying the text responses, then the computer has passed a typical version of what we refer to today as the "Turing test."

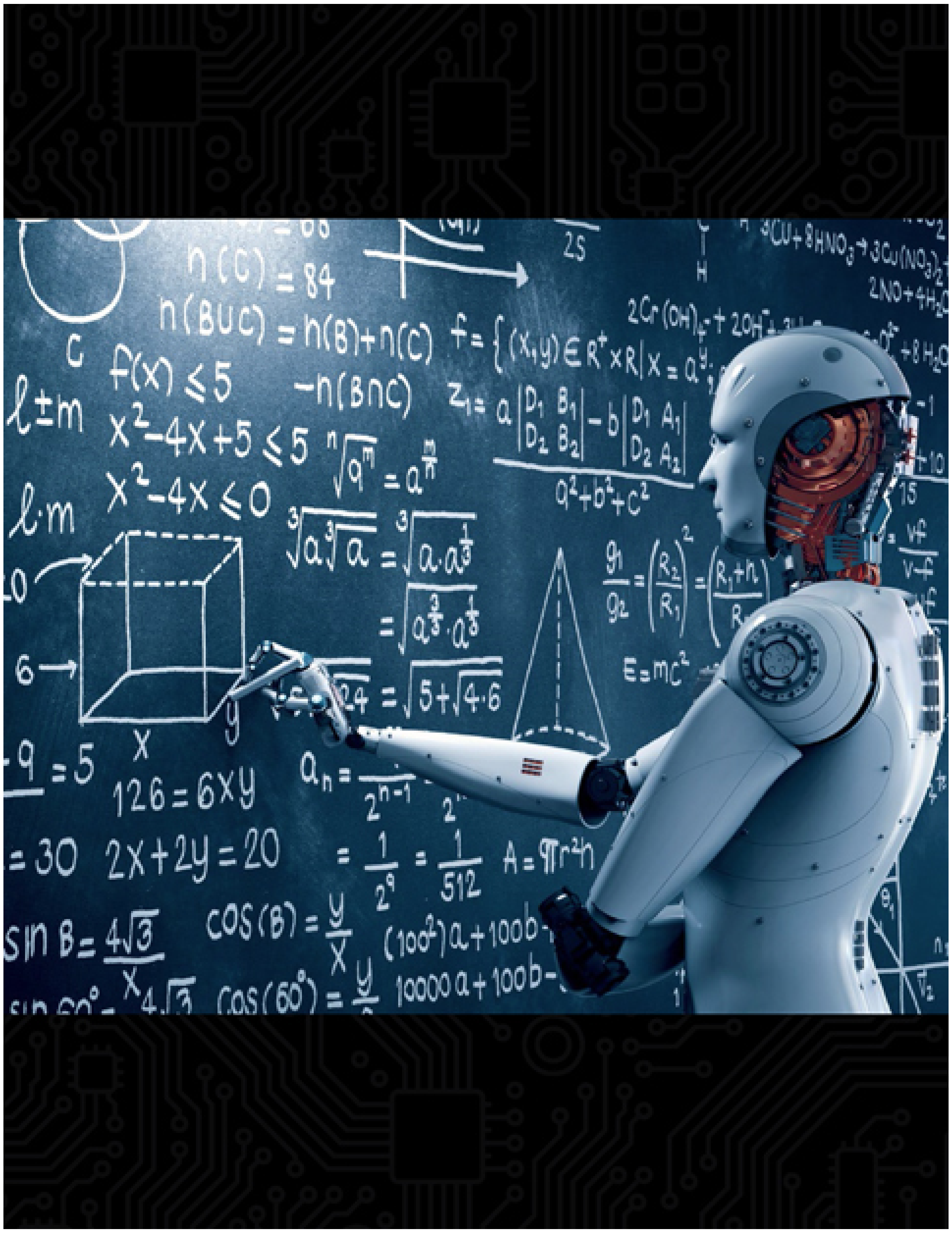
Today, the Loebner competition is held each year, in part, to honor the computer programmers who create programs that come closest to passing the Turing test. Of course, the Turing test has stimulated much debate and controversy over the years. For example, if the computer is actually far more "intelligent" than humans, it will need to pretend to be less intelligent because the test is focused on imitating humans. As a result, devious and funny techniques that introduce

typing errors, change the topic of conversation, inject jokes, ask the judges questions, and so forth were often used to fool judges. In 2014, a conversational robot developed by programmers in Russia passed a version of the test, pretending to be a thirteen-year-old Ukrainian boy named Eugene Goostman.

Another challenge to the value of the Turing test is that the level of expertise of the human judges can easily change the outcome of the test. However, whatever we think about the test's ability to detect "intelligence," it definitely inspires creativity in computer programmers and engineers.

SEE ALSO "Darwin among the Machines" (1863), *Giant Brains, or Machines That Think* (1949), Natural Language Processing (1954), ELIZA Psychotherapist (1964), Chinese Room (1980), Moravec's Paradox (1988)

Norbert Wiener wrote that the machine that “can learn and can make decisions on the basis of its learning, will in no way be obliged to make such decisions as we should have made, or will be acceptable to us.”



1950

THE HUMAN USE OF HUMAN BEINGS

Norbert Wiener (1894–1964), the influential American mathematician and philosopher, was one of the key originators of the field of *cybernetics*, which is concerned with feedback in many realms of human endeavor and technology. According to AI expert Daniel Crevier, Wiener believed that feedback mechanisms are "information-processing devices: they receive information and then make a decision based on it. Wiener speculated that all intelligent behavior is the consequence of feedback mechanisms; perhaps by definition, intelligence is the outcome of receiving and processing information."

In his book *The Human Use of Human Beings* (1950), Wiener contemplates ways in which humans and machines will cooperate, and his vision certainly has applicability today with people in nearly constant electrical communication: "It is the thesis of this book that society can only be understood through a study of the messages and the communication facilities which belong to it; and that in the future, development of these messages and communication facilities, messages between man and machines, between machines and man, and between machine and machine, are destined to play an ever increasing part."

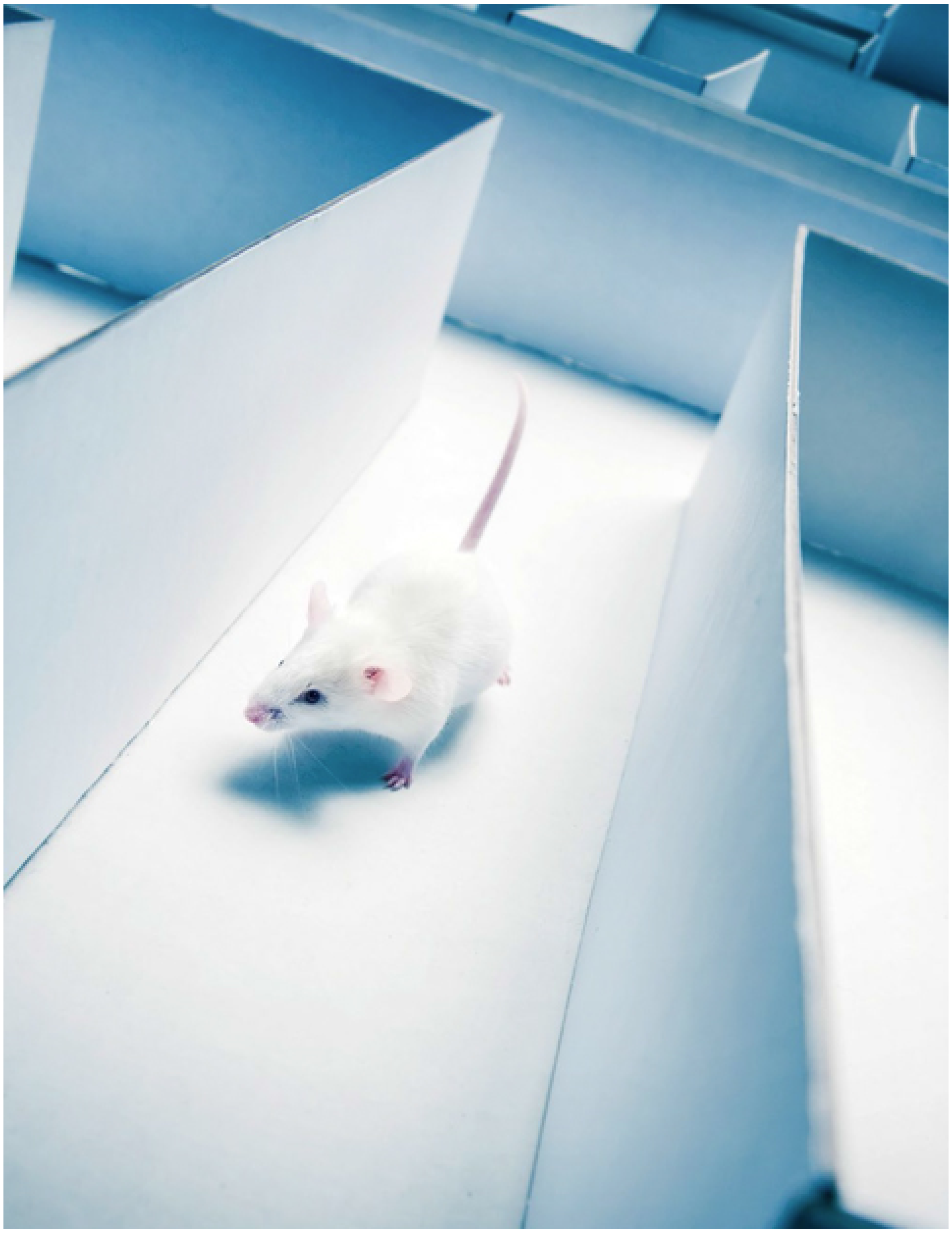
He presciently contemplated the need for future machines that learn, but he also provided warnings about delegating decision-making processes to literal-minded machines: "Any machine

constructed for the purpose of making decisions, if it does not possess the power of learning, will be completely literal minded. Woe to us if we let it decide our conduct, unless we have previously examined its laws of action, and know fully that its conduct will be carried out on principles acceptable to us! [The machine that] can learn and can make decisions on the basis of its learning, will in no way be obliged to make such decisions as we should have made, or will be acceptable to us. [If a man gives responsibility to the machine], whether it can learn or not, is to cast his responsibility to the winds, and to find it coming back seated on the whirlwind."

These words of caution have relevance for us today, with many futurists warning of the need for formulating paths to safe artificial general intelligence.

SEE ALSO "Darwin among the Machines" (1863), Tesla's "Borrowed Mind" (1898), Dartmouth AI Workshop (1956), Intelligence Explosion (1965), Deep Learning (1965)

Reinforcement learning is an approach for teaching useful actions to software agents so as to maximize a cumulative reward. Famous early uses included maze solving and learning to play checkers, tic-tac-toe, and backgammon.



1951

REINFORCEMENT LEARNING

Reinforcement learning is reminiscent of a simple behavior observed in cats seeking rewards. In the early 1900s, the psychologist Edward Thorndike (1874–1949) placed cats inside boxes from which they could escape only by stepping on a switch. After some wandering, the cats would eventually step on the switch by chance, and the door would open, and the cat would be rewarded, for example, with food. After the cats learned to associate this behavior with the reward, they escaped with increasing speed, up to, eventually, a maximum rate of escape.

In 1951, cognitive scientist Marvin Minsky (1927–2016) and his student Dean Edmunds built SNARC (Stochastic Neural Analog Reinforcement Calculator), a neural network machine consisting of 3,000 vacuum tubes to simulate 40 connected neurons. Minsky used the machine to study a scenario involving a simulated rat navigating a maze. When, by chance, the rat made a sequence of useful moves and escaped from a maze, the connections corresponding to these moves were strengthened, thus reinforcing the desired behavior, facilitating learning. Other famous early examples of reinforcement learning devices include systems for playing checkers (1959), tic-tac-toe (1960), and backgammon (1992).

As implied by these examples, in its simplest definition, reinforcement learning (RL) is an area of machine learning that involves traversing a collection of states to receive a reward or to

maximize a cumulative reward. The “learner” discovers which actions yield the highest reward by repeatedly testing actions. Today, RL is often combined with deep learning, which involves a large simulated neural network, often to recognize patterns in data. With RL, a system or machine can learn without explicit instructions; thus, machines like self-driving cars, industrial robots, and drones can develop and improve skills through trial and error and experience. One practical challenge of broadly applying RL is that it requires a large amount of data and practice simulations.

SEE ALSO Tic-Tac-Toe (c. 1300 BCE), Artificial Neural Networks (1943), Machine Learning (1959), Backgammon Champion Defeated (1979), Checkers and AI (1994)

IBM's Shoebox machine listened as an operator spoke numbers and arithmetic commands such as "Five plus three plus eight minus nine. Total."

1952

SPEECH RECOGNITION

A recent issue of *The Economist magazine* likened today's speech-recognition devices to "casting a magic spell," allowing people to "control the world through words alone." This reminds us of novelist Arthur C. Clarke's assertion that any sufficiently advanced technology is indistinguishable from magic. "The fast-emerging technology of voice computing proves Clarke's point. . . . Say a few words into the air, and a nearby device can grant your wish."

The science and technology of *speech recognition*, which enables the recognition of spoken language by machines, has had a long history. In 1952, Bell Laboratories developed the AUDREY system, which made use of vacuum-tube circuitry and understood spoken numerical digits. Ten years later, IBM's Shoebox machine at the 1962 Seattle World's Fair understood sixteen words, including the digits *0* through *9*, and it would perform arithmetic operations if it heard words like *plus*. In 1987, the "Julie" doll, offered by the US toy company Worlds of Wonder, could understand some simple phrases and reply.

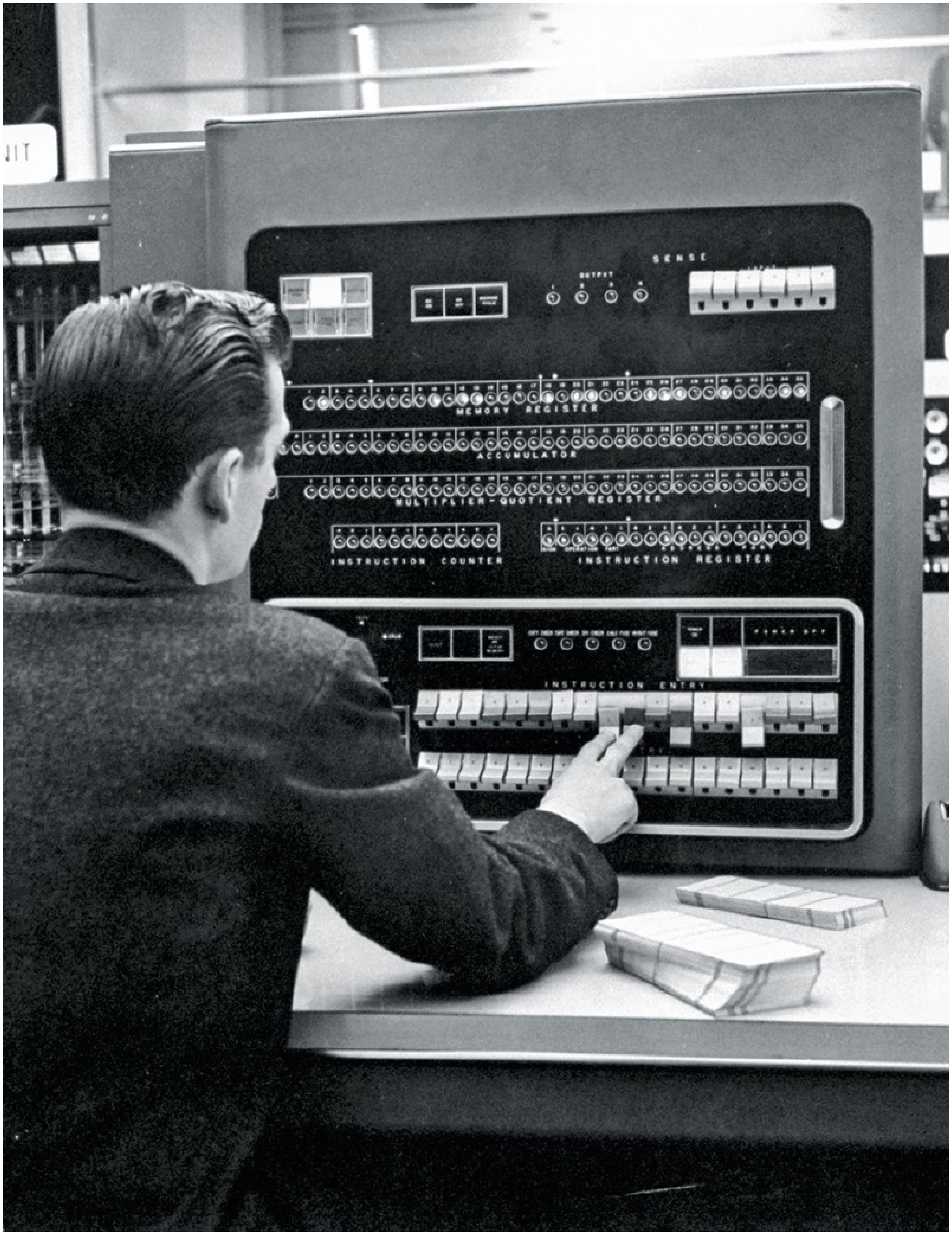
The technology enabling machines to recognize speech has evolved considerably. Historically, it has involved the hidden Markov model (HMM), a statistical method for predicting whether a sound corresponds to a word, and today recognition makes significant use of deep learning (i.e., artificial neural nets with multiple layers) to achieve high accuracy. For example, a speech-recognition system

may hear a stream of sound in a noisy environment and make a number of “guesses” about what is being said by determining the probability of various words and phrases it has encountered in training text. Special software applications may have knowledge of the likelihood of a particular phrase being used, and decide, for example, whether “abdominal aortic aneurysm” should be ranked high or low, depending upon whether it is heard by radiological dictation software or if it is said in a car that is awaiting a simple command.

Of course, today, countless digital assistants—in our homes, cars, offices, and mobile phones—respond to voice questions and commands, and they can also help with dictation of notes. Blind persons and the physically disabled may also benefit from speech input.

SEE ALSO Speech Synthesis (1939), Artificial Neural Networks (1943), Natural Language Processing (1954)

In 1954, in a famous public demonstration of a research project known as the "Georgetown-IBM experiment," Russian was automatically translated into English by an "electronic brain" in the form of an IBM 701 computer, shown here.


SENSE
MEMORY REGISTER
ACCUMULATOR
MULTIPLIER-QUOTIENT REGISTER
INSTRUCTION COUNTER
INSTRUCTION REGISTER
INSTRUCTION ENTRY

1954

NATURAL LANGUAGE PROCESSING

In 1954, an IBM press release proclaimed: “Russian was translated into English by an electronic ‘brain’ today for the first time. . . . The famous 701 computer . . . within a few seconds, turned the sentences into easily readable English. A girl who didn’t understand a word of the language of the Soviets punched out the Russian messages on IBM cards.” The press release continued by explaining that “the ‘brain’ dashed off its English translations on an automatic printer at the breakneck speed of two and a half lines per second.”

In 1971, computer scientist Terry Winograd (b. 1946) wrote SHRDLU, a program that translated human commands such as “Move the red block next to the blue pyramid” into physical actions. Today, *natural language processing* (NLP) often involves many AI subfields, including speech recognition, natural language understanding (e.g., machine reading comprehension), and speech synthesis. One of the goals is to facilitate natural interactions between humans and computers.

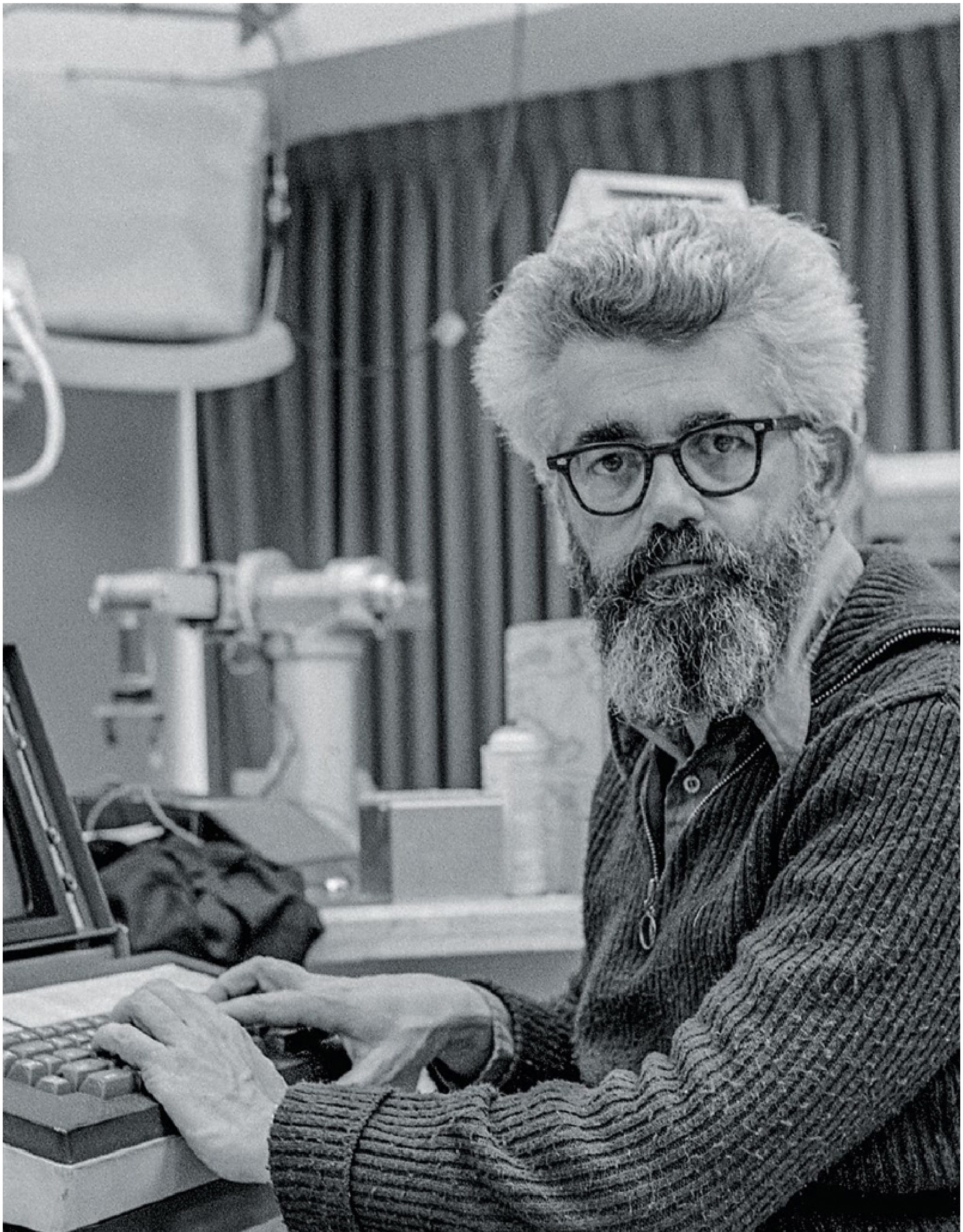
The early days of NLP usually involved the use of complicated sets of manually created rules; but in the 1980s, NLP made increasing use of machine learning algorithms that learned rules through analysis of large sets of example language input. Typical NLP tasks may involve machine translations (e.g., translating Russian into English), question answering (e.g., “What is the capital of France?”), sentiment analysis (emotions and attitudes on a topic),

etc. Analyzing input from text, audio, and video, NLP is employed in diverse realms, including spam-filtering for email, summarizing information in long articles, and question-answering by smartphone apps.

NLP is challenging for many reasons. For example, in speech recognition, the sounds of adjacent words blend into one another, and the computing system must also take into account syntax (i.e., grammar), semantics (i.e., meaning), and pragmatics (i.e., purpose or goal)—along with the many classes of ambiguity in language where words take on different meanings in different contexts. Today, significant use of artificial neural network methods helps to improve accuracy.

SEE ALSO Speech Synthesis (1939), Artificial Neural Networks (1943), Turing Test (1950), Speech Recognition (1952), Machine Learning (1959), Licklider's "Man-Computer Symbiosis" (1960), ELIZA Psychotherapist (1964), SHRDLU (1971), Paranoid PARRY (1972), Watson on *Jeopardy!* (2011)

The Dartmouth Summer Research Project on Artificial Intelligence is considered to be an important event in the history of AI and where the term *artificial intelligence*, selected by computer scientist John McCarthy (pictured here in 1974), started to gain acceptance.



1956

DARTMOUTH AI WORKSHOP

"In the summer of 1956," writes journalist Luke Dormehl, "when Elvis Presley was scandalizing audiences with his hip gyrations . . . and President Dwight Eisenhower authorized 'In God we trust' as the US national motto—AI's first official conference took place." That conference—the Dartmouth Summer Research Project on Artificial Intelligence—is where the term *artificial intelligence*, coined by computer scientist John McCarthy (1927–2011), started to gain acceptance.

The workshop was formally proposed by McCarthy of Dartmouth College, Marvin Minsky (1927–2016) of Harvard University, Nathaniel Rochester (1919–2001) of IBM, and Claude Shannon (1916–2001) of Bell Telephone Laboratories, as follows: "We propose that a 2-month, 10-man study of artificial intelligence be carried out during the summer. . . . [We conjecture] that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made . . . if a carefully selected group of scientists work on it together for a summer." The proposal also made specific mention of several other key areas of consideration, including "neuron nets" and "randomness and creativity."

During the meeting, Carnegie Mellon's Allen Newell (1927–1992) and Herbert Simon (1916–2001) presented their Logic Theorist, a program for proving theorems in symbolic logic. Author Pamela McCorduck writes about the Dartmouth Workshop: "They had in common a belief . . . that what we call thinking could indeed take place outside the human cranium, that it could be understood in a formal and scientific way, and that the best nonhuman instrument for doing it was the digital computer."

Partly because of the complexity of AI technology, and because meeting participants came and went on different dates, the expectations for the conference were probably a bit too high. Nevertheless, the Dartmouth AI workshop brought together a diverse group of researchers who influenced the field for the next twenty years.

SEE ALSO Artificial Neural Networks (1943), Natural Language Processing (1954), Licklider's "Man-Computer Symbiosis" (1960)

The first version of the perceptron was implemented in software on an IBM 704 computer, pictured here in a 1957 photo. The IBM 704 was among the first mass-produced computers with floating-point arithmetic hardware.


IBM

1957

PERCEPTRON

Today, artificial neural networks (ANNs) are used in countless applications, including pattern recognition (e.g., face recognition), time-series prediction (e.g., predicting whether a stock price will rise), signal processing (e.g., filtering out noise), and more. The basics of neural networks are discussed on page 77, and one historically important step on the ladder to fully functional ANNs were *perceptrons*, developed in 1957 by psychologist Frank Rosenblatt (1928–1971). In 1958, based partly on Rosenblatt's enthusiasm, the *New York Times* proclaimed that the perceptron was "the embryo of an electronic computer that [the Navy] expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence."

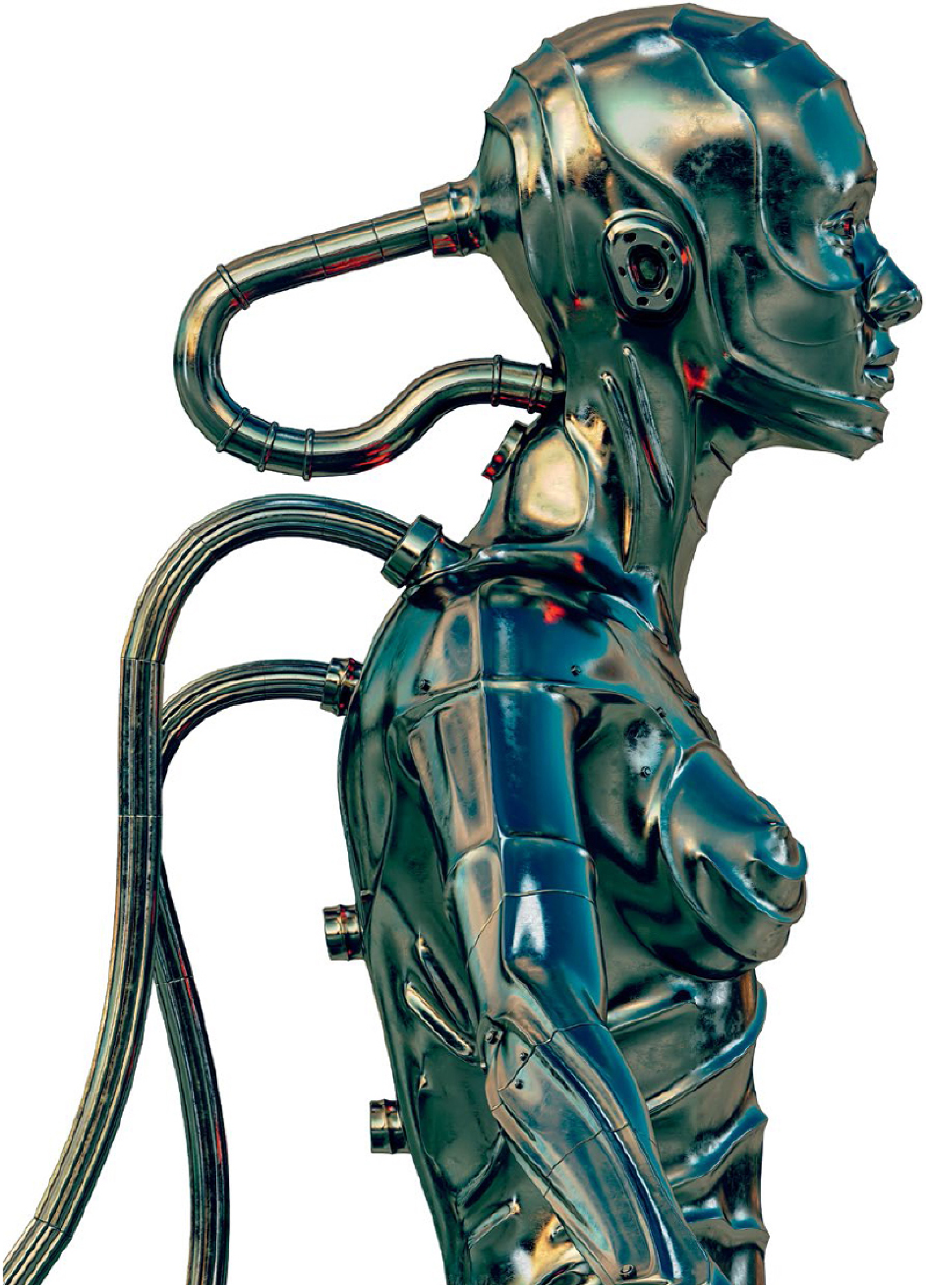
The perceptron originally had three levels of connected "neurons" (i.e., simple computational units). The first level was a 20 × 20 array of photocells corresponding to the retina of an eye. The second level contained connector cells that received input from the photocells, the initial connections of which were random. The third level consisted of output cells that labeled the object placed in front of the machine (e.g., "triangle"). If the perceptron guessed correctly (or incorrectly), researchers might strengthen (or weaken) the electrical connections between the cells that led to the label.

The first version was implemented in software on an IBM 704 computer. A second version, the Mark 1 perceptron, was implemented in special hardware and was a trainable machine that

could learn to classify certain patterns by modifying the strength of connections between the neurons. Mathematical weights were actually encoded in potentio-meters, and changes in weights during the learning process were performed by electric motors. It was hoped that the device could perform a wide range of pattern-recognition tasks; but, alas, the hype and hope exceeded the limitations of such a simple model. In fact, the 1969 book *Perceptrons* by MIT's Marvin Minsky (1927–2016) and Seymour Papert (1928–2016) explicitly demonstrated the limitations of simple perceptrons and dampened interest in this nascent field of machine learning. Later on, however, it became much clearer that configurations of neurons with more layers could have immeasurable value and applications.

SEE ALSO Artificial Neural Networks (1943), Machine Learning (1959), Deep Learning (1965)

American scientist and economist Yoshihiro Francis Fukuyama (b. 1952) has called transhumanism—which usually involves the use of technology to enhance human, mental, and physical capacities—the world's most dangerous idea.



1957

TRANSHUMANISM

“The coming of artificial intelligence will likely be the most significant event in the history of the human species,” writes transhumanist philosopher Zoltan Istvan. “The key, of course, is not to let artificial intelligence run wild and out of sight, but to already be cyborgs and part machines ourselves, so that we can plug right into it wherever it leads.”

In *New Bottles for New Wine*, published in 1957, biologist Julian Huxley (1887–1975) coined the term *transhumanism* when he suggested that “the human species can . . . transcend itself . . . by realizing new possibilities of and for his human nature. . . . The human species will be on the threshold of a new kind of existence, as different from ours as ours is from that of Peking man. It will at last be consciously fulfilling its real destiny.”

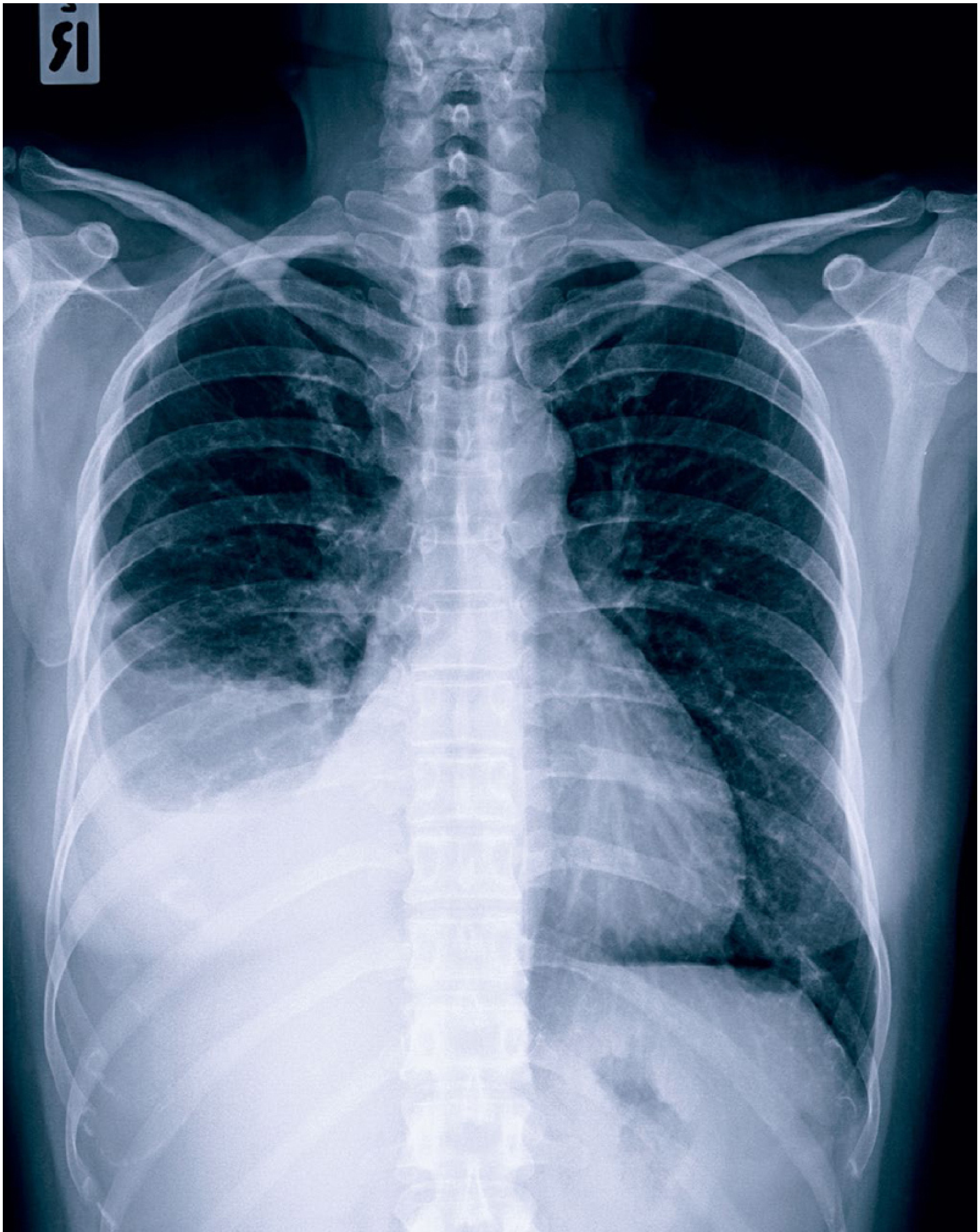
The modern idea of transhumanism, espoused by philosopher-futurist Max More (b. 1964) and many others, usually involves the use of technology to enhance human mental and physical capacities. The idea is that perhaps one day we will become “posthuman,” or even immortal through genetic manipulation, robotics, nanotechnology, computers, or mind uploading to virtual worlds—and because we will fully understand the biological basis of aging. Already we are getting glimmers of how to use brain-computer interfaces to link us to advanced artificial intelligences to expand our

cognitive abilities, and the closer we get to fully understanding the biological basis of aging, the closer we may get to immortality.

If your body or mind could survive indefinitely, would "you" actually persist? All of us are changed by our experiences—and these changes are usually gradual, which means that you are nearly the same person you were a year ago. However, if your normal or enhanced body survived continuously for a thousand years, gradual mental changes would accumulate, and perhaps an entirely different person would eventually inhabit the body. The thousand-year-old person might be nothing like you. "You" would no longer exist. There would be no moment of death at which you had ceased to exist, but you would slowly fade away over the millennia, like a sand castle being transformed by an ocean of time.

SEE ALSO The Consciousness Mill (1714), Searches for the Soul (1907), Licklider's "Man-Computer Symbiosis" (1960), Living in a Simulation (1967), Spielberg's *A.I. Artificial Intelligence* (2001)

In 2017, Stanford researchers developed a machine-learning algorithm that could outperform radiologists at diagnosing pneumonia. Shown here is a chest X-ray indicating right pleural effusion.



1959

MACHINE LEARNING

Artificial intelligence expert Arthur Lee Samuel (1901–1990) is credited as being among the first to use the term *machine learning*, which gained prominence in his 1959 paper "Some Studies in Machine Learning Using the Game of Checkers," published in the *IBM Journal of Research and Development*. In the paper, he explained that "programming computers to learn from experience" may eventually eliminate much of the need for explicit, task-specific programming and instructions.

Today, one of the major computing methods and enablers of artificial intelligence is machine learning. It plays a role in computer vision, speech understanding, autonomous robots, self-driving cars, face recognition, email filtering, optical character recognition, providing product recommendations, identifying likely cancers, detecting data breaches, and more. Many forms of machine learning require large sets of input data for training, to help make predictions and classifications.

With *supervised machine learning*, algorithms may be fed data samples that are labeled with information, so that the system can later make predictions when subsequent data is provided in an unlabeled fashion. For example, consider a system that ingests a collection of 100,000 images of lions and tigers that humans have correctly labeled as "lion" or "tiger." Subsequently, the supervised learning algorithm should be able to distinguish lions from tigers it

has not seen before. With *unsupervised machine learning*, unlabeled data is employed, so that the system may discover hidden patterns. For example, the system may determine that thirty-year-old women who stop purchasing albacore canned tuna may be pregnant and, thus, are targets for baby-product ads.

Note that machine-learning approaches can be fallible—for example, if the input data is biased, incorrect, or even manipulated in a malicious way. We should be careful not to overrely on certain automated approaches when making decisions on who should qualify for loans, be offered jobs, or receive parole. This principle applies in countless areas of machine-mediated decision making.

SEE ALSO Artificial Neural Networks (1943), Reinforcement Learning (1951), Natural Language Processing (1954), Knowledge Representation and Reasoning (1959), Deep Learning (1965), Genetic Algorithms (1975), Swarm Intelligence (1986), Adversarial Patches (2018)

MYCIN was an expert system that used AI to identify bacteria causing severe infections, such as meningitis, and to recommend treatments. MYCIN employed a simple inference engine, and a knowledge base of around 600 rules. Shown here are *Streptococcus pneumoniae* bacteria that can cause meningitis.

1959

KNOWLEDGE REPRESENTATION AND REASONING

“For a system to be intelligent,” writes computer scientist Nils Nilsson, “it must have knowledge about its world and the means to draw conclusions from, or at least act on, that knowledge. Humans and machines alike therefore must have ways to represent this needed knowledge in internal structures, whether encoded in protein or silicon.” Nowadays, much of the attention on AI seems to be on machine learning and statistical algorithms for applications such as image recognition. Nevertheless, logic-based knowledge representation and reasoning (KR) still has a big role to play in many areas.

KR is the field of AI concerned with representing information in such a way that computer systems can efficiently use the information to make medical diagnoses and legal recommendations, as well as to facilitate intelligent dialog systems such as Siri on the iPhone or Alexa on the Amazon Echo. As just one example, a *semantic network* is sometimes used as a form of KR to represent semantic (i.e., meaning) relationships between concepts. These semantic networks often take the form of graphs, with vertices representing concepts and edges (i.e., connecting lines) indicating the semantic relationships between them. KR also has application in automatic reasoning, including the automated proving of mathematical theorems.

Some of the early work in AI KR includes the General Problem Solver, a computer program developed in 1959 by Allen Newell (1927–1992), Herbert Simon (1916–2001), and colleagues to analyze goals and solve simple general problems (e.g., the Tower of Hanoi). Later, the Cyc project—started in 1984 by Douglas Lenat (b. 1950)—employed numerous analysts documenting various areas of commonsense reasoning to help AI systems perform human-like reasoning (e.g., the Cyc inference engine employs logical deductions and inductive reasoning). Today, AI researchers in the area of KR deal with many issues, including ensuring that the knowledge base can be updated as needed to allow efficient development of new inferences. The researchers are also concerned with how uncertainty and vagueness can best be addressed in KR systems.

SEE ALSO Aristotle's *Organon* (c. 350 BCE), Tower of Hanoi (1883), Perceptron (1957), Machine Learning (1959), Expert Systems (1965), Fuzzy Logic (1965)

Joseph Licklider wrote: “The hope is that . . . human brains and computing machines will be coupled together very tightly, and that the resulting partnership will think as no human brain has ever thought.”

1960

LICKLIDER'S "MAN-COMPUTER SYMBIOSIS"

In 1960, psychologist and computer scientist Joseph Licklider (1915–1990) published a seminal paper titled "Man-Computer Symbiosis." He begins his essay by explaining a symbiotic relationship of a fig tree, which is pollinated by a *Blastophaga* wasp whose eggs and larvae derive nourishment from the tree. In the same way, Licklider proposed, humans and computers could form a symbiotic relationship. In the early years of symbiosis, humans would set the goals and formulate the hypotheses as computers prepared the way for insights. Some problems, he wrote, "simply cannot be formulated without computing-machine aid."

Rather than envisioning computer-based entities replacing humans, Licklider was more aligned with Norbert Wiener (1894–1964), whose theories of cybernetics tended to focus on close interactions between humans and machines. In the paper, he explains: "The hope is that . . . human brains and computing machines will be coupled together very tightly, and that the resulting partnership will think as no human brain has ever thought and process data in a way not approached by the information-handling machines we know today."

Licklider also discusses "thinking centers" that he believed would incorporate the functions of traditional libraries, and he suggested the need for natural-language processing for symbiosis.

In his essay, Licklider concedes that “electronic or chemical ‘machines’ will outdo the human brain in most of the functions we now consider exclusively within its province,” and he provides examples of chess playing, problem solving, pattern recognizing, and theorem proving. He clarifies that “the computer will serve as a statistical-inference, decision-theory, or game-theory machine to make elementary evaluations of suggested courses of action. . . . Finally, it will do as much diagnosis, pattern-matching, and relevance-recognizing as it profitably can. . . .”

Nearly sixty years later, Licklider’s paper still raises important questions about the potential union of human intelligence and AI: When the day comes that we have a coupling with machines even more than we have today, will a symbiotic person still be considered a “human”? Will such a person ever consider disengaging from the computer?

SEE ALSO “Darwin among the Machines” (1863), Natural Language Processing (1954), Transhumanism (1957)

Portrait of Eliza Doolittle, a seller of flowers from George Bernard Shaw's play *Pygmalion*, by artist William Bruce Ellis Ranken (1881–1941). ELIZA was named after Miss Doolittle, since Doolittle could convincingly simulate a sophisticated and educated person, partly by having improved her language skills.

1964

ELIZA PSYCHOTHERAPIST

ELIZA is a computer program that responds to natural-language input (e.g., typed text), simulating a conversation between the user and a psychotherapist. Developed in 1964 by computer scientist Joseph Weizenbaum (1923–2008), the program became famous as one of the first and most convincing "chatterbots" (i.e., conversational simulators). In fact, Weizenbaum was shocked and distressed at the degree to which some people revealed deep emotions and personal information during their dialogues with ELIZA, as if they considered ELIZA to be a real human being with the capacity for empathy.

ELIZA was named after Eliza Doolittle from the Irish playwright George Bernard Shaw's 1912 comedy *Pygmalion*. In the play, Professor Henry Higgins teaches Eliza, an uneducated woman, how to speak properly and, thus, convincingly mimic an upper-class lady. Similarly, Weizenbaum's ELIZA was programmed to respond to certain keywords and phrases, providing the illusion of real human empathy. Some researchers believed that the program could actually help individuals suffering from certain psychological conditions.

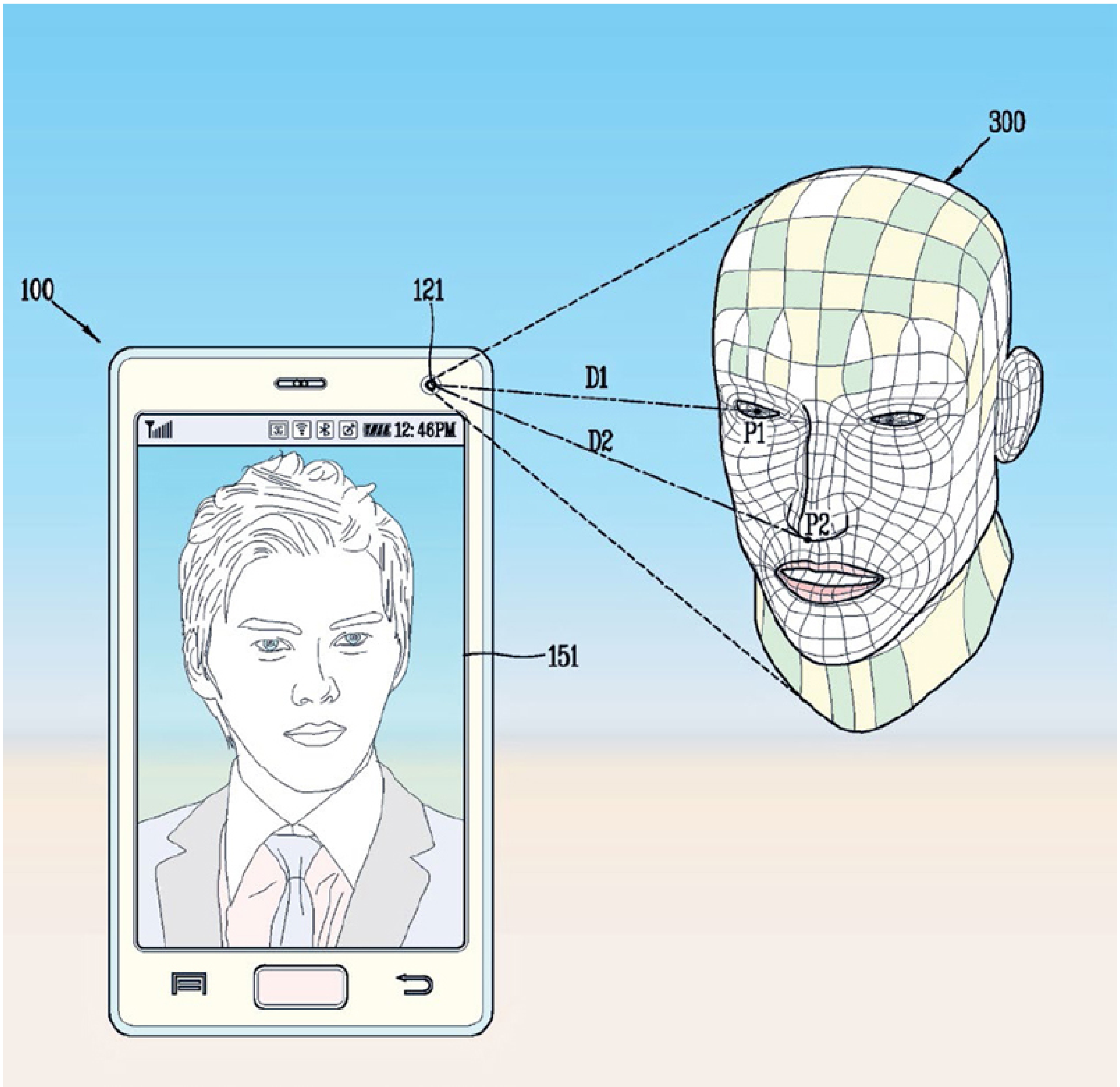
As Weizenbaum watched people interacting with ELIZA, he grew alarmed at their growing dependence on computers and at how people could be deceived. In his landmark 1966 technical paper on ELIZA, he wrote: "[In] artificial intelligence . . . machines are made to behave in wondrous ways, often sufficient to dazzle even the most

experienced observer. But once a particular program is unmasked, once its inner workings are explained . . . its magic crumbles away; it stands revealed as a mere collection of procedures. . . . The observer says to himself 'I could have written that.' With that thought, he moves the program in question from the shelf marked 'intelligent,' to that reserved for curios. . . . The object of this paper is to cause just such a re-evaluation of the program about to be 'explained.' Few programs ever needed it more."

Today, chatbots are often used in dialog systems for customer service and for various forms of online virtual assistance and mental-health therapy. They are also used in some toys, or to assist consumers shopping online, or as advertising agents.

SEE ALSO Turing Test (1950), Natural Language Processing (1954), Paranoid PARRY (1972), Ethics of AI (1976)

Figure from US patent 9,703,939 for a method of securely unlocking (accessing) a mobile phone using the phone's camera and face recognition.



1964

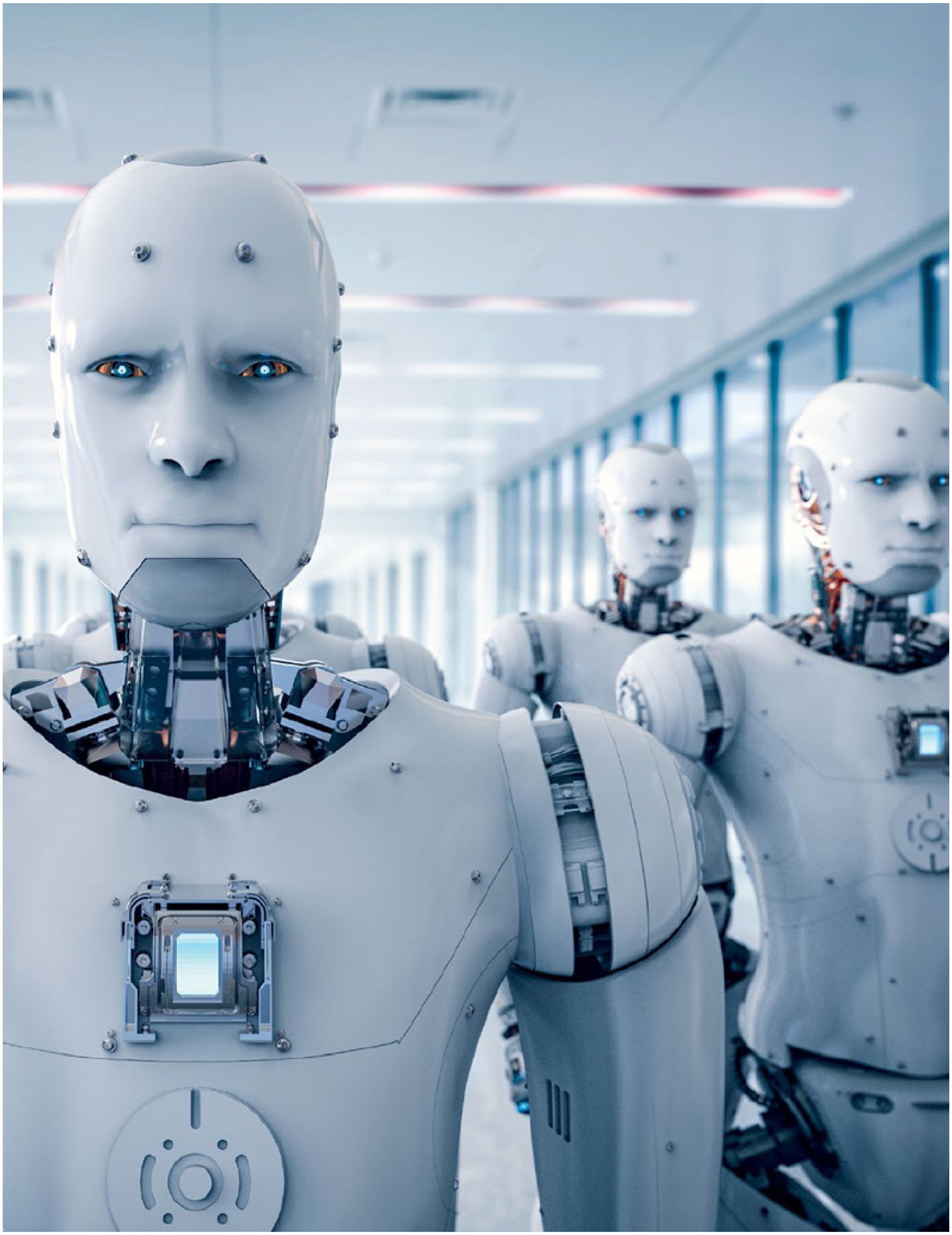
FACE RECOGNITION

Face recognition systems attempt to identify people from images or video footage, often by comparing facial features (e.g., relative position of the eyes and nose) with those in an image database. Some modern systems use 3-D sensors to capture information and to increase accuracy when confronted with variations in lighting and viewing angles, and certain smartphones use infrared light to illuminate a user's face during the authentication process. Many challenges exist for accurate face recognition, such as when people are wearing accessories like hats and sunglasses, or even makeup, but today, in some circumstances, algorithms outperform humans in recognizing faces. The origin of face recognition "technology" could be traced to nineteenth-century England, where a regular system of prison photography was introduced in 1852, as a more humane method than branding, to keep track of prisoners and to share with other police departments when they escaped. One of the pioneers of more advanced face recognition was mathematician and computer scientist Woody Bledsoe (1921–1995), who worked on earlier forms of face recognition in 1964. At the time, he noted that the task was particularly difficult due to head rotation and tilt, lighting, facial expression, age, etc. Bledsoe and other early pioneers tended to rely on significant human collaboration with the computer, as humans manually extracted image facial coordinates from photos by hand using a graphics tablet (e.g., drawing pad).

Over the years, face recognition systems have employed a variety of techniques, including eigenfaces, hidden Markov models, and dynamic link matching. As outlined by technologist Jesse Davis West, face recognition has several important applications today: “Law enforcement agencies are using face recognition to keep communities safer. Retailers are preventing crime and violence. Airports are improving travelers’ convenience and security. And mobile phone companies are using face recognition to provide consumers with new layers of biometric security.” Nevertheless, one may wonder if this could mark a worrisome turning point in civilization, when people are no longer able to go out in public and remain anonymous.

SEE ALSO Optical Character Recognition (OCR) (1913), Speech Recognition (1952), AIBO Robot (1999)

In 1965, Irving J. Good expressed concerns regarding a potential superhuman "intelligence explosion" in which AIs designed increasingly better versions of themselves.



1965

INTELLIGENCE EXPLOSION

In 1965, British mathematician Irving J. Good (1916–2009)—who worked as a cryptologist with fellow computer scientist Alan Turing—expressed concerns regarding a potential superhuman “intelligence explosion” in a paper titled “Speculations Concerning the First Ultraintelligent Machine.” In the paper, Good writes: “Let an ultraintelligent machine be defined as a machine that can far surpass all the intellectual activities of any man however clever. Since the design of machines is one of these intellectual activities, an ultraintelligent machine could design even better machines; there would then unquestionably be an intelligence explosion, and the intelligence of man would be left far behind. Thus, the first ultraintelligent machine is the last invention that man need ever make, provided that the machine is docile enough to tell us how to keep it under control.”

In other words, if humans built an AGI (i.e., artificial general intelligence that is not limited to a restricted field of knowledge and capability), it might be self-improving with engineering capabilities that allowed it to recursively redesign its hardware and software. As just one example configuration, perhaps such AGIs could use neural nets and evolutionary algorithms to build hundreds of separate modules that intercommunicate and collaborate with increasing sophistication, speed, and efficiency. Attempts to confine or isolate the potentially dangerous AI from the rest of the Internet might fail.

Even if programmed with a beneficent goal and task, such as manufacturing better lightbulbs, what if it decided to convert the entire North American continent into a lightbulb manufacturing facility?

Of course, there may be many reasons why such a superintelligence would be unlikely, such as the need to rely on slow humans and hardware networks. On the other hand, it is possible that an intelligence explosion would also immensely benefit humanity in the race to cure diseases and solve environmental problems. But what would be the sociological effects of superintelligent weapons—or even artificial companions that exhibited an intellect and (simulated) empathy that is greater than our human spouses?

SEE ALSO “Darwin among the Machines” (1863), Lethal Military Robots (1942), *The Human Use of Human Beings* (1950), Leakproof “AI Box” (1993), Paperclip Maximizer Catastrophe (2003), “Call Them Artificial Aliens” (2015)

AI expert systems are often created by extracting the specialized knowledge of human experts, artistically represented by glimmering lightbulbs in this drawing. The expert information is converted into a set of probabilistic rules.

1965

EXPERT SYSTEMS

According to journalist Luke Dormehl, AI “expert systems” are “attempts to create clones of flesh-and-blood human experts . . . by extracting their specialized knowledge and turning it into a set of probabilistic rules.” In a best-case scenario, expert systems could, in principle, be used to cram the know-how of an expert gastroenterologist, financial adviser, or lawyer into a computerized device, and have the resultant AI systems give useful advice to all.

Expert systems started to be explored in the 1960s and made use of a knowledge base (containing representations of facts and rules) and inference engines (to apply the rules and perform evaluations). Rules might be of an “if-then” form, such as “If a patient with a particular demographic exhibits a particular symptom, then there is a certain probability that he or she has a particular condition.”

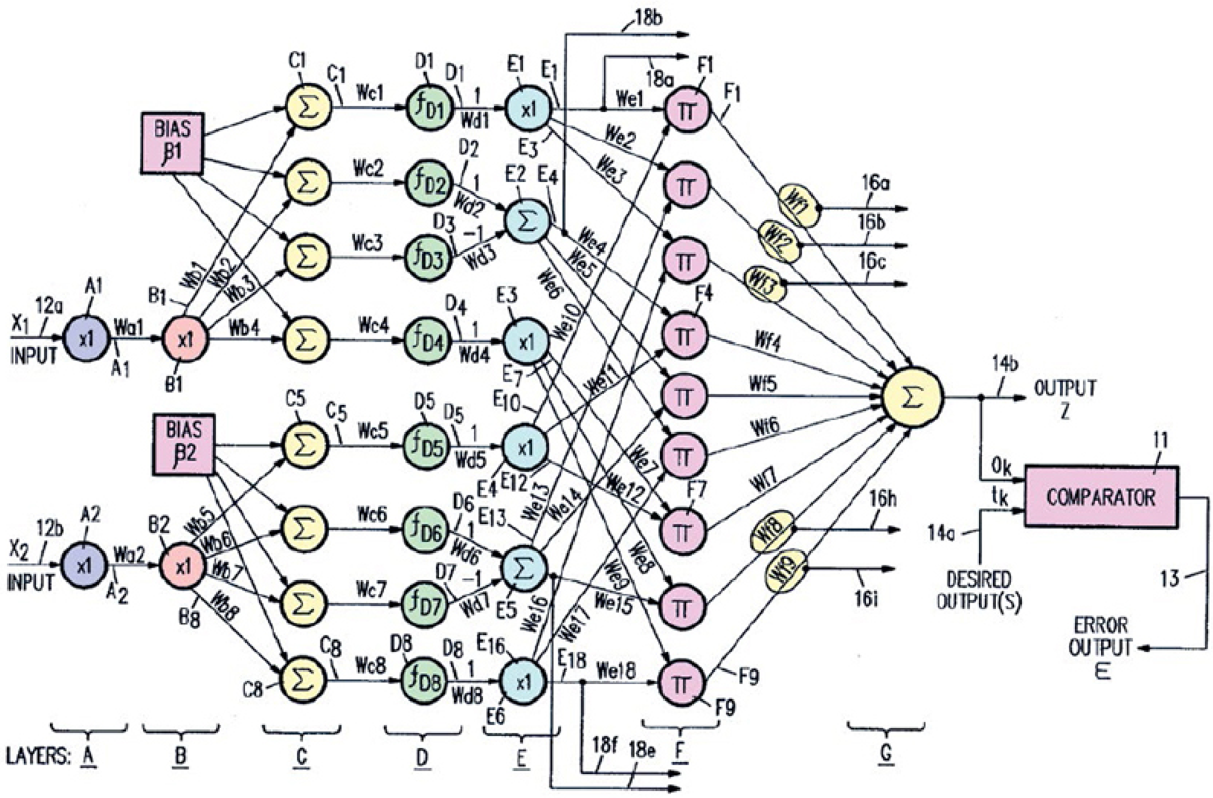
Applications of expert systems may broadly include diagnosis, prediction, planning, classifying, and related areas involving specialized domains of expertise, in areas ranging from medicine to evaluating insurance risks or potential locations for mineral exploration. Useful expert systems also often have inference engines that can provide explanations, so that users can understand the chain of reasoning. An example of a famous early expert system is Dendral (short for Dendritic Algorithm), a Stanford University project started in 1965 to help chemists identify unknown organic molecules based on information from mass spectra. Another famous early

example is MYCIN, a Stanford University AI system developed in the 1970s to help diagnose bacterial infections and recommend antibiotics and dosages. Early expert systems were often coded in LISP or Prolog.

One of the challenges of expert systems often involves acquiring and codifying knowledge from busy experts in particular fields, or from books and papers. It can also be challenging to organize knowledge into a collection of facts and rules that experts agree on, along with various numerical weights applied (to signify likelihood or importance). Today, many people use “recommender systems,” a somewhat related area of AI that is more focused on predicting user preferences in areas ranging from movies and books to financial services and potential marriage partners.

SEE ALSO *The Human Use of Human Beings* (1950), Knowledge Representation and Reasoning (1959), Deep Learning (1965)

Diagram from US patent 5,579,439 that provides a fuzzy logic design for an intelligent controller in a plant-control system. The design includes an artificial neural network for generating fuzzy logic rules and membership functions data. “The fuzzification layer of the learning mechanism neural network can be constructed of four layers A, B, C, D of neurons.”



1965

FUZZY LOGIC

“Fuzzy set theory has been used in commercial applications of expert systems and control devices for trains and elevators,” writes scientist Jacoby Carter. “It has also been combined with neural nets to control the manufacture of semiconductors. By incorporating fuzzy logic and fuzzy sets in production systems, significant improvements have been gained in many AI systems. This approach has been particularly successful with ambiguous data sets or when the rules are imperfectly known.”

Classical two-valued logic is concerned with conditions that are either true or false. Fuzzy set theory, which focuses on members of a set that have *degrees* of membership, was introduced by mathematician and computer scientist Lotfi Zadeh (1921–2017) in 1965, and in 1973 he provided the details of *fuzzy logic* (FL). Classical two-valued logic is concerned with conditions that are either true or false, whereas FL allows a continuous range of truth values.

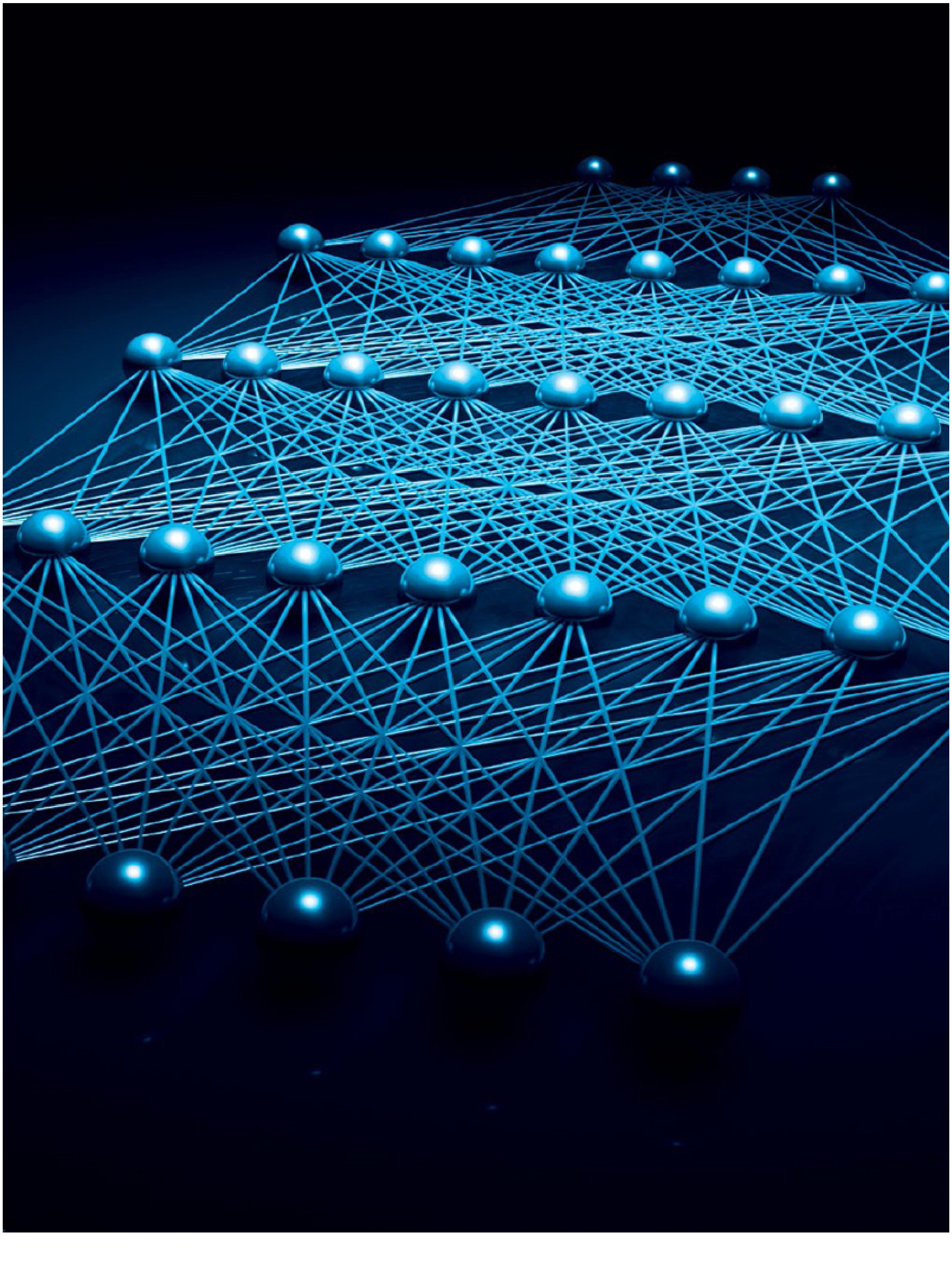
FL has a wide range of practical applications. As an example, consider a temperature-monitoring system for a device. A membership function may exist for the concepts cold, warm, and hot, but a single measurement may consist of three values such as “not cold,” “slightly warm,” and “slightly hot.” Zadeh believed that if feedback controllers could be programmed to make use of

imprecise, noisy input, they could be more effective and easier to implement.

One milestone in the history of fuzzy logic occurred in 1974, when Ebrahim Mamdani (1942–2010) of the University of London used it to control a steam engine. In 1980, FL was used to control a cement kiln. Various Japanese companies have used FL to control water-purification processes and train systems. FL has also since been used to control steel mills, self-focusing cameras, washing machines, fermentation processes, automobile-engine controls, anti-lock braking systems, color-film developing systems, glass processing, computer programs used in financial trading, and systems used for recognizing subtle differences in written and spoken languages.

SEE ALSO Aristotle's *Organon* (c. 350 BCE), Boolean Algebra (1854), Expert Systems (1965)

Deep neural nets (DNNs) contain multiple intermediate layers of artificial neuron units (e.g., ranging from just a few layers to dozens of layers), which increase the ability of DNNs to learn. DNNs comprise the “architecture” upon which deep learning takes place.



1965

DEEP LEARNING

Artificial intelligence involves approaches that enable machines to mimic human intelligence. Machine learning (see page 99) is a form of AI that enables machines to improve at some task through practice and experience. *Deep learning* is a form of machine learning that allows systems to train themselves to perform tasks, like playing a game or recognizing a cat in a photo, using deep neural nets (DNNs), which have multiple intermediate layers of artificial neuron units, in contrast to shallow nets, which employ just a few layers. Although the phrase *deep learning* was not introduced until 1986, Soviet mathematician Alexey Ivakhnenko (1913–2007) conducted early work, in the form of supervised deep multilayer perceptrons, in 1965.

Generally speaking, the multiple layers of neurons can perform feature extraction on the data at different levels of a hierarchy (e.g., responding to simple edges at one level and to facial features at another). Deep learning may involve *backpropagation*, a process in which the system can pass information in a reverse direction, from output to input, in order to teach the system when it has made errors so as to improve the results.

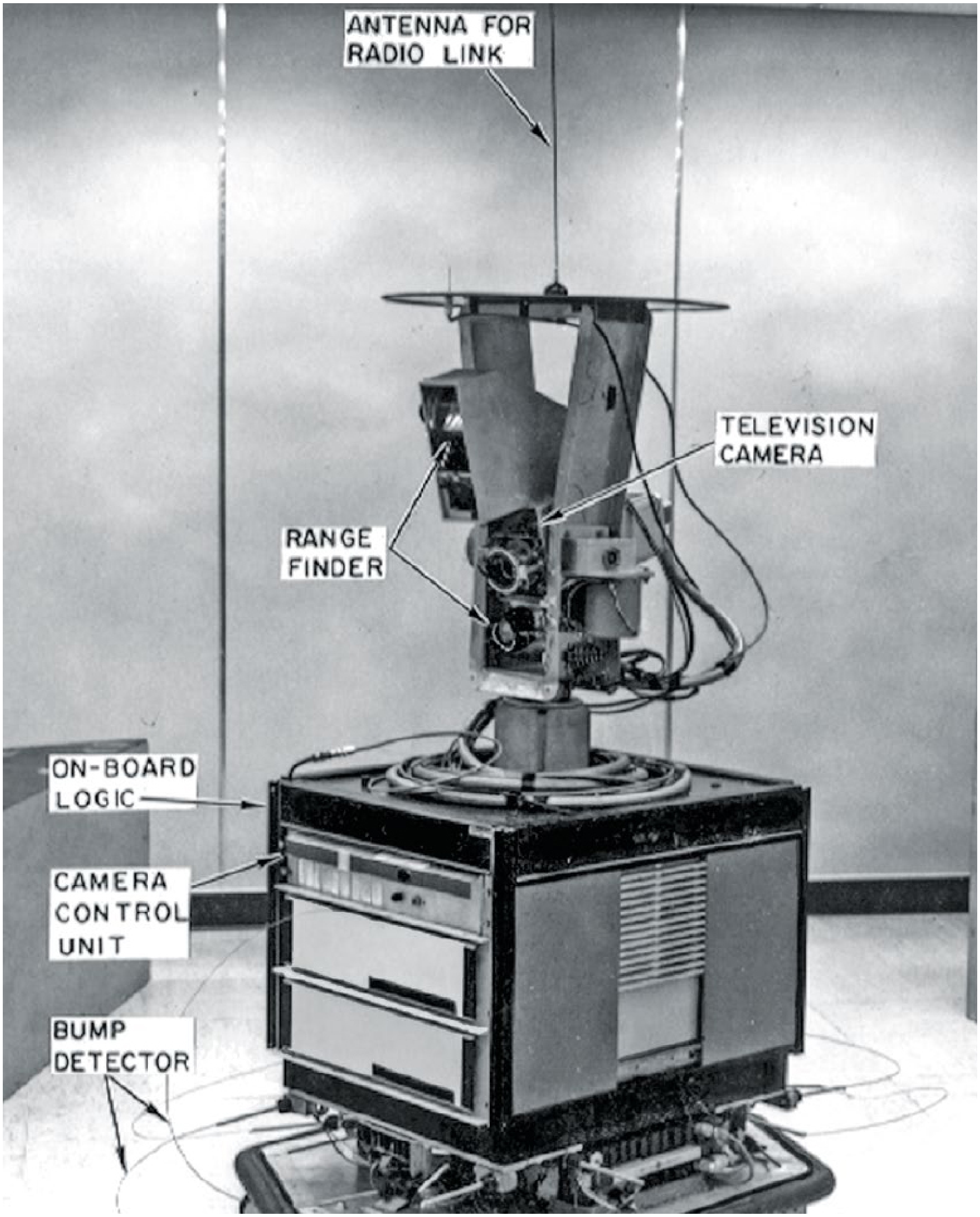
Deep learning has been successfully applied to speech recognition, computer vision, natural language processing, social networking, human language translations, drug design, identifying specific style periods for paintings, making product

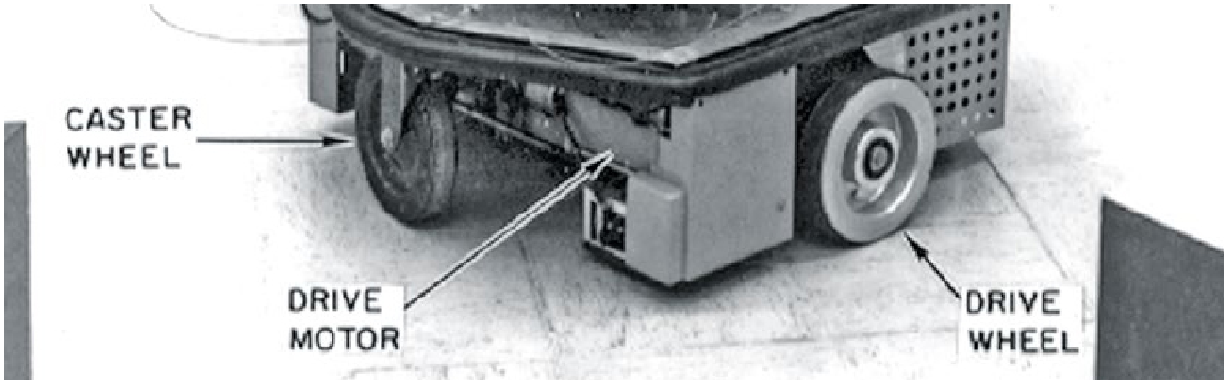
recommendations, estimating the value of different marketing actions, performing image restoration and cleaning, playing games, identifying people in photos, and more.

As technologist Jeremy Fain writes: “Ultimately, deep learning has pushed machine learning across a threshold. Whereas machine learning had some success in automating repetitive tasks or data analytics, it is now bringing the future to life in the form of computers that can see, hear and play all types of games.”

SEE ALSO Artificial Neural Networks (1943), Reinforcement Learning (1951), Perceptron (1957), Machine Learning (1959), Computer Art and DeepDream (2015), Adversarial Patches (2018)

Shakey was notable for setting serious milestones in the fields of pattern recognition and computer vision, problem solving, natural-language processing, and information representation. Shakey contained a TV camera, distance rangefinders, collision sensors, and motors for steering.




CASTER
WHEEL
DRIVE
MOTOR
DRIVE
WHEEL

1966

SHAKEY THE ROBOT

In 1970, *Life* magazine referred to Shakey as the world's "first electronic person" who could soon "travel about the Moon for months at a time without a single beep of direction from Earth." Although the report of this fascinating robot was quite an exaggeration, Shakey was notable for setting serious milestones in the fields of pattern recognition and computer vision, problem solving, natural-language processing, and information representation.

Developed between 1966 and 1972 at the Stanford Research Institute, Shakey was among the earliest serious attempts to design a general-purpose, autonomous robot that could move around, perceive its surroundings, monitor the execution of its plans, and reason about its actions. The project was funded by the US Defense Advanced Research Projects Agency (DARPA) and was primarily programmed using LISP. In order for the robot to perform well, its world was limited to several rooms, connected by corridors, with doors, light switches, and objects that Shakey could push around. An operator typed in commands like "Push the block off the platform," and Shakey attempted to explore and identify the platform, push a ramp to the platform, drive up the platform, and push the block off.

Shakey relied on various levels of programs. For example, one level used routines for route planning, motor control, and the capturing of sensory information, while a middle level was concerned with actually moving to a designated position and processing the

images from Shakey's TV camera. High-level programming involved task planning and performing sequences of actions to achieve goals.

Not surprisingly, Shakey the robot got its name from its jerky motion as it drove about. Carrying around an antenna for a radio and video link to a DEC PDP computer, Shakey also contained a TV camera, distance rangefinders, collision sensors, and motors for steering. The development of Shakey led to important research in AI, including the development of search algorithms for pathfinding, and feature extraction methods in computer vision.

SEE ALSO Natural Language Processing (1954), SHRDLU (1971), AIBO Robot (1999), ASIMO and Friends (2000), AI on Mars (2015)

As computers become more powerful, perhaps someday we will be able to simulate entire worlds—fanciful and realistic—and reality itself. It is possible that more advanced beings are already doing this somewhere in the universe.


456t
fast
783 7785655434
76590-3245
7-875/6453

1967

LIVING IN A SIMULATION

"Our universe seems real, but is it really?" writes author Jason Koebler. "As humans get better at simulating artificial intelligence, it seems at least plausible that we could create life that is [conscious]. And if we can create conscious life, who's to say that the universe, as we know it, wasn't created by superintelligent artificial intelligence . . . ?"

Could we be living in a computer simulation and be artificial intelligences ourselves? The hypothesis that the universe is a digital computer was pioneered by German engineer Konrad Zuse (1910–1995) in 1967. Other researchers, including Ed Fredkin (b. 1934), Stephen Wolfram (b. 1959), and Max Tegmark (b. 1967), have suggested that the physical universe may be running on a cellular automaton or discrete computing machinery—or be a purely mathematical construct.

In our own small pocket of the universe, we have already developed computers with the ability to simulate lifelike behaviors using software and mathematical rules. One day we may create thinking beings that live in simulated spaces as complex and vibrant as a rain forest. Perhaps we'll be able to simulate reality itself, and it is possible that more advanced beings are already doing this elsewhere in the universe.

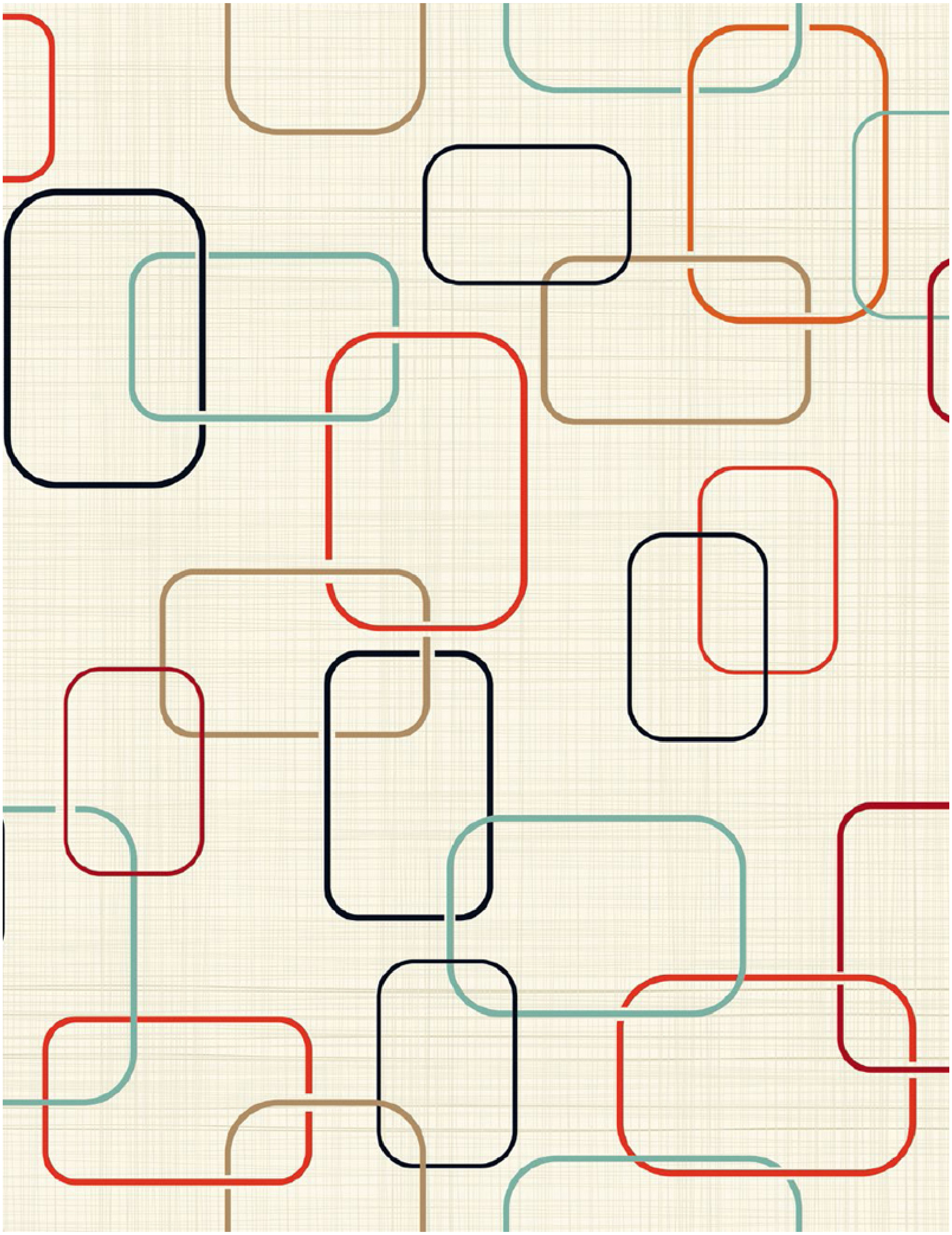
What if the number of these simulations is larger than the number of universes? Astrophysicist Martin Rees (b. 1942) suggests that if

this were the case, ". . . as they would if one universe contained many computers making many simulations," then it is likely that *we* are artificial life. Rees continues, "Once you accept the idea of the multiverse . . . , in some of those universes there will be the potential to simulate parts of themselves, . . . and we don't know what our place is in this grand ensemble of universes and simulated universes."

Physicist Paul Davies (b. 1946) expanded on the notion of a multiverse with multiple simulated realities in a 2003 *New York Times* article: "Eventually, entire virtual worlds will be created inside computers, their conscious inhabitants unaware that they are the simulated products of somebody else's technology. For every original world, there will be a stupendous number of available virtual worlds—some of which would even include machines simulating virtual worlds of their own. . . ."

SEE ALSO The Consciousness Mill (1714), Searches for the Soul (1907), Artificial Life (1986), "Call Them Artificial Aliens" (2015)

Cybernetic Serendipity was notable for demonstrating numerous aspects of computer-aided creativity. The book also featured computer designs made on mechanical computer plotters.



1968

CYBERNETIC SERENDIPITY

Cybernetic Serendipity: The Computer and the Arts (1968) was a groundbreaking illustrated book and catalogue for the well-attended exhibition at the Institute of Contemporary Arts in London (and later in Washington, DC, and San Francisco). Edited and curated by British art critic Jasia Reichardt (b. 1933), the book and simultaneous exhibit were notable for demonstrating numerous aspects of computer-aided creativity, from visual art to music, poetry, storytelling, dance, animation, and sculpture, inspiring a generation of experimental collaborations involving artists, scientists, and engineers.

In 1948, American mathematician and philosopher Norbert Wiener (1894–1964) defined *cybernetics* as "the scientific study of control and communication in the animal and the machine"; today the term has taken on a broader meaning involving the control of numerous kinds of systems using technology such as electronic devices. *Serendipity* usually refers to unexpected, delightful, and useful outcomes made by accident and chance.

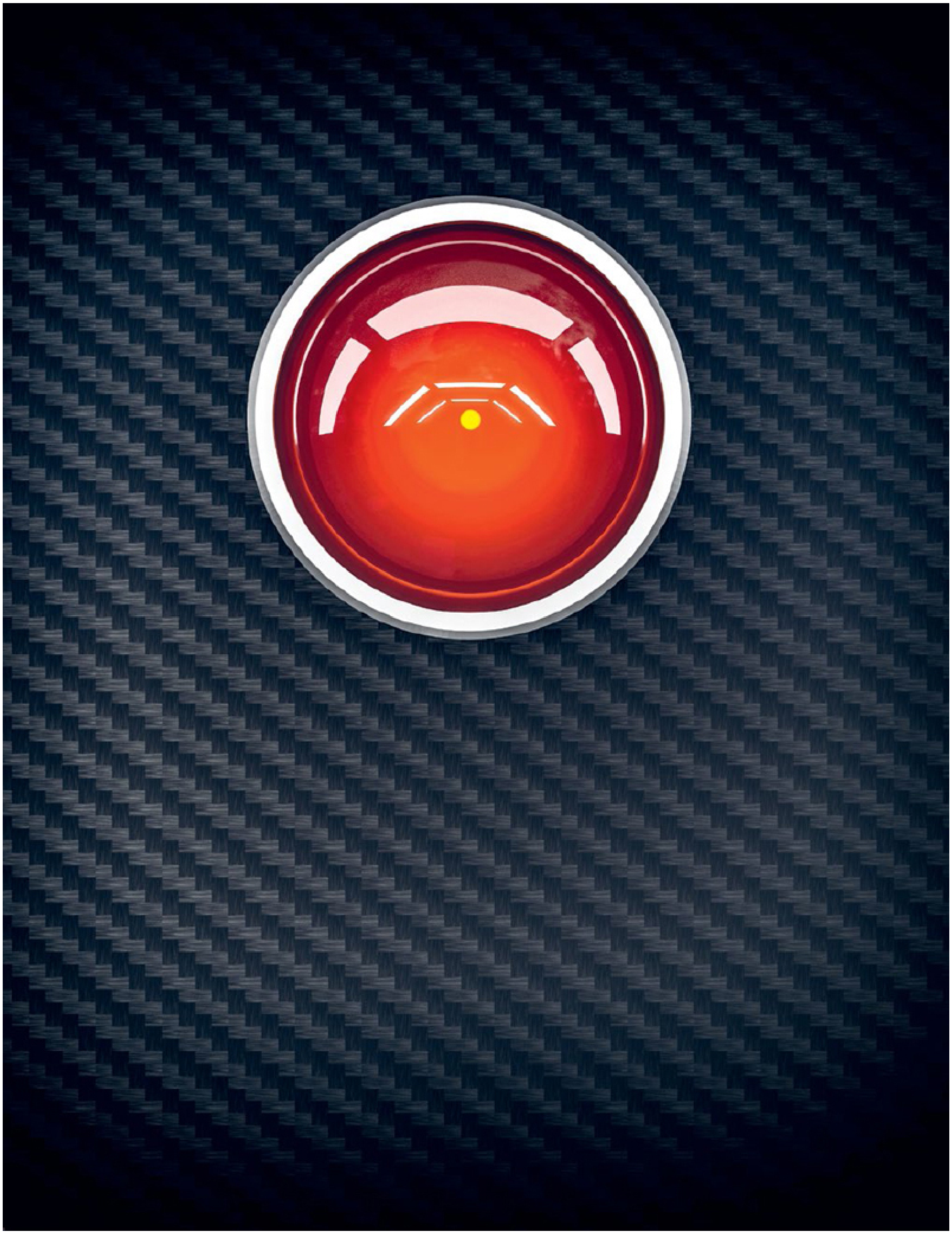
With contributors that included artists, composers, poets, and computer programmers, *Cybernetic Serendipity* raised questions about the nature of art and chance. Some of the most stimulating parts of the book are beautiful computerized Japanese haiku poems, accompanied by algorithmic recipes to generate the poems. For example, one haunting poem begins: "Eons deep in the ice, I paint

all time in a whorl. . . .” Another section concerns the generation of “Little Grey Rabbit-type stories” that are constructed by a computer using simple rules. A typical story might begin like this: “The sun shone over the woods. Across the fields softly drifted the breeze while then the clouds, which calmly floated all afternoon, moved across the fields . . . ”

In addition to poetry and stories, *Cybernetic Serendipity* featured painting machines and a range of computer art made on mechanical computer plotters or CRT displays, inspiring people interested in the nascent field of algorithmic and/or generative art. The book also included musical scores generated from traces of the New York City skyline or splatters of ink, various forms of electronic music machines, computer-programmed choreography, sound-activated mobiles, pendulum drawing machines and harmonographs, architectural works, pseudo-Mondrian paintings, and more.

SEE ALSO Ramon Llull’s *Ars Magna* (c. 1305), Lagado Book-Writing Engine (1726), Computational Creativity (1821), Computer Art and DeepDream (2015)

Artist's rendering of HAL 9000's famous red all-seeing camera eye from *2001: A Space Odyssey.*



1968

HAL 9000

“My mission responsibilities range over the entire operation of the ship, so I am constantly occupied,” explains HAL 9000, the fictional AI in the famous 1968 movie *2001: A Space Odyssey*. “I am putting myself to the fullest possible use, which is all I think that any conscious entity can ever hope to do.” Unfortunately for the crew, HAL later becomes a killer who must be terminated.

HAL is important partly because a number of prominent AI researchers say they were inspired to explore the field of AI after watching the movie, the screenplay of which was written by Stanley Kubrick (1928–1999) and Arthur C. Clarke (1917–2008). Interestingly, HAL embodies many of the capabilities we expect today of future artificial general intelligences (AGIs) that can behave intelligently across a wide range of goals and environments. This sentient machine is capable of computer vision, speech recognition, face recognition, speech output, natural-language processing, chess playing, and various forms of advanced reasoning, planning, and problem solving. HAL can even read lips, appreciate art, exhibit emotions, strive for self-preservation, and interpret the emotions of humans.

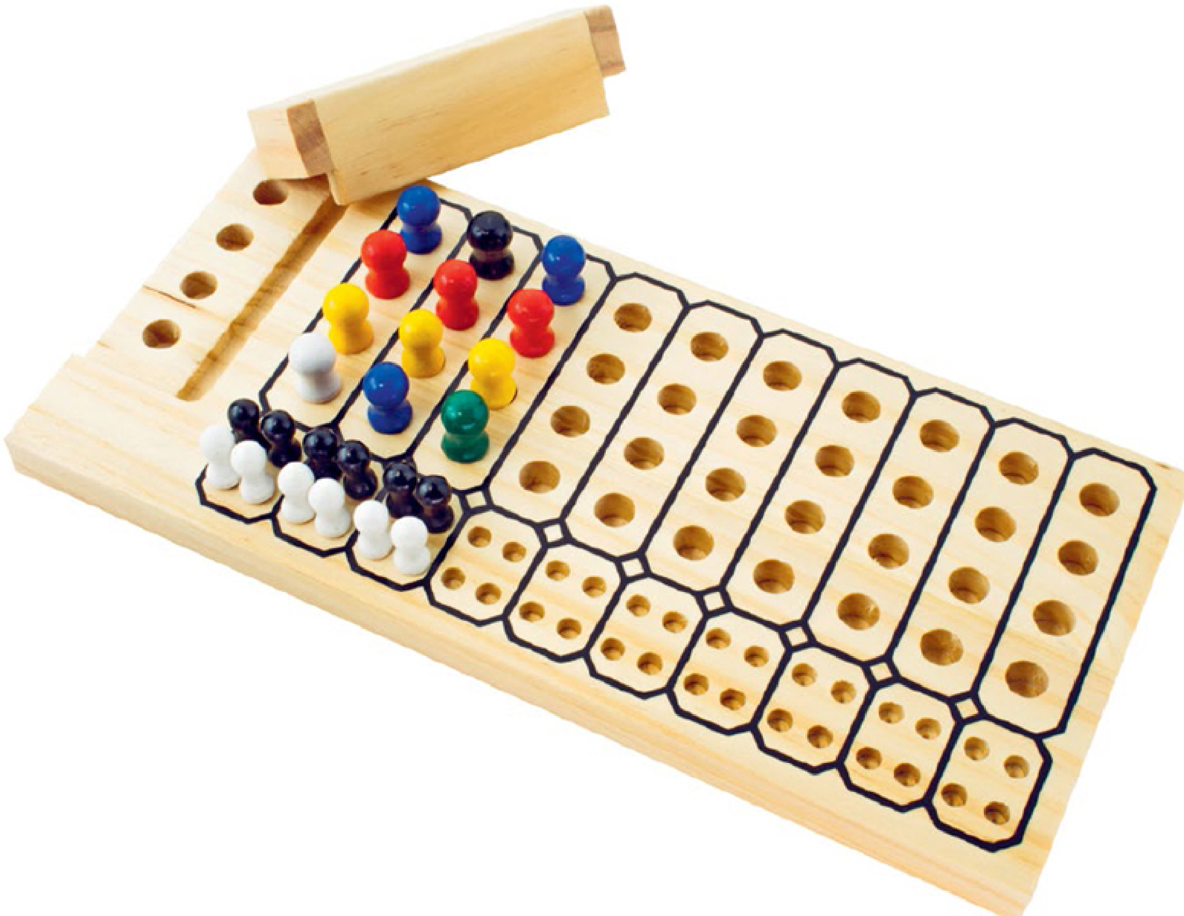
In the 1960s, when the film was being made, experts had predicted that by 2001 AIs like HAL would be possible. Even though AI researcher Marvin Minsky (1927–2016) was an adviser on the film

set, it is apparent today that many more years are needed to create a HAL-like entity, with his full range of capabilities.

In a haunting scene from the movie, after HAL becomes dangerous and must be disabled, we see an astronaut slowly removing computer modules as HAL continually degrades and slowly sings. Perhaps one lesson of HAL today is that we may not be fully aware of the effects of pervasive AI use from childhood to death. We would be eminently wise to consider both the pros and cons of pervasive AI in the future, as these entities become more powerful and we rely on them for countless forms of decision making.

SEE ALSO Natural Language Processing (1954), *Colossus: The Forbin Project* (1970), *The Terminator* (1984)

Mastermind playing board. The “decoding” region of the board is where colored pegs are placed. The initial secret sequence of colored pegs would be placed at left, covered by a shield. The smaller pegs at bottom track the correctness of guesses.



1970

MASTERMIND

Mastermind®, a colorful code-breaking board game, has been the subject of passionate AI research for decades. Invented in 1970 by Mordecai Meirowitz (b. 1930), an Israeli postmaster and telecommunications expert, the game was initially rejected by all mainstream game companies. Nevertheless, the game went on to sell over 50 million copies, making it the most successful new game of the 1970s.

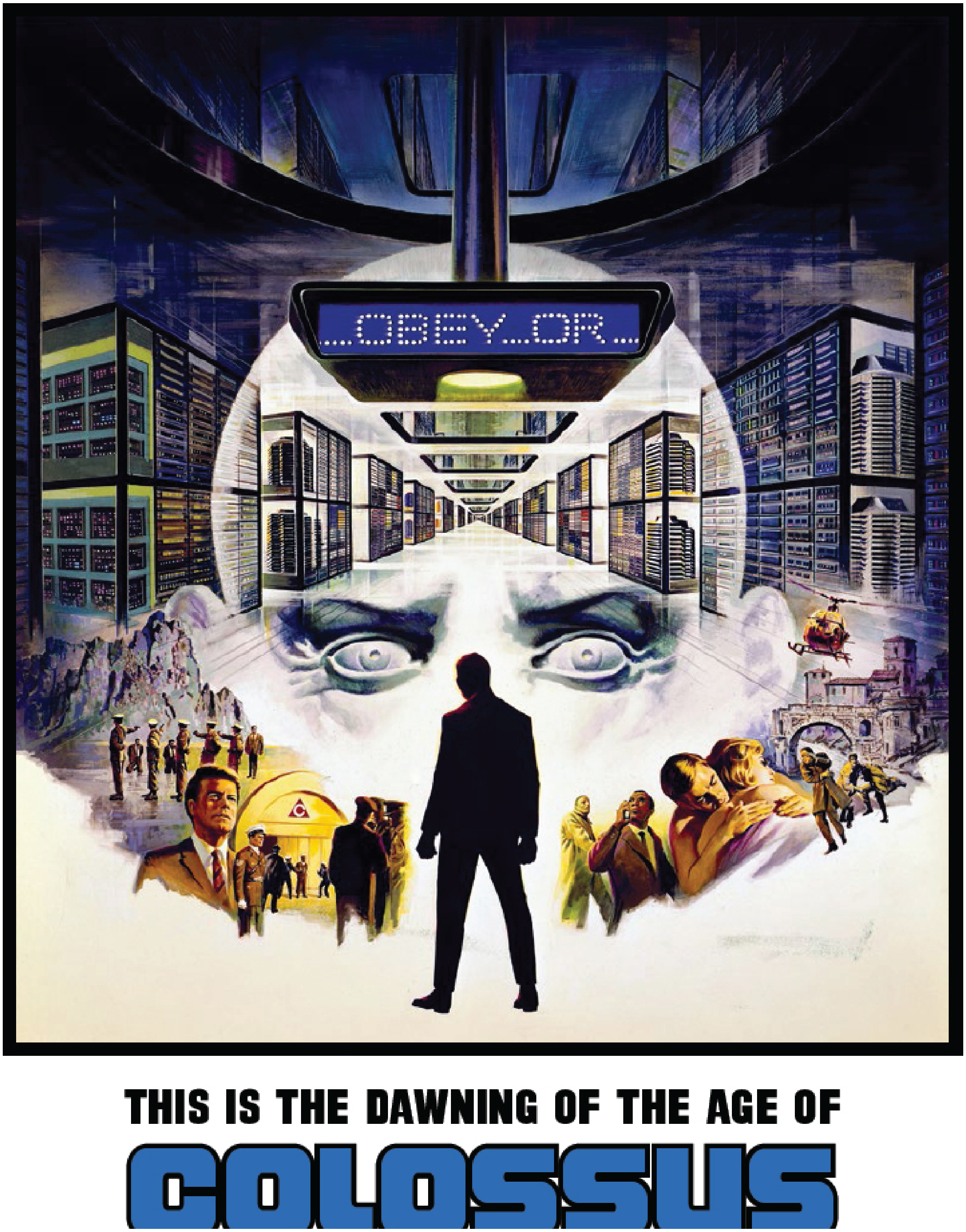
To play the game, a code-maker selects a sequence of four colors, represented by colored pegs that come in six different colors. The opponent must guess the code-maker's secret sequence—with as few guesses as possible. Each guess is presented in the form of a sequence of four colored pegs. The code-maker reveals how many of those pegs are both the correct color and in the correct position and how many more are the correct color but in the wrong position. For example, the secret code may be *green-white-blue-red*. The guess may be *orange-yellow-blue-white*. In this situation, the code-maker indicates that the player has one peg of the correct color in the correct position and one peg of the correct color in the wrong position, but he doesn't mention the specific color names. The game continues with more guesses. A code-maker selects from a possible 6^4 (or 1,296) possible combinations, assuming there are six colors and four positions.

In 1977, American computer scientist Donald Knuth (b. 1938) published a strategy that enables a player to always guess the correct code within five guesses. This was the first known algorithm to solve Mastermind, and numerous papers followed. In 1993, Kenji Koyama and Tony W. Lai published a strategy with a maximum of six guesses required in the worst case, but with an average number of guesses of only 4.340. In 1996, Zhixiang Chen and colleagues generalized previous results to the case of *n* colors and *m* positions.

Mastermind has also been studied several times using genetic algorithms inspired by evolutionary biology. In 2017, researchers from Asia University in Taiwan employed AI strategies of reinforcement learning and obtained an average number of guesses of 4.294.

SEE ALSO Tic-Tac-Toe (c. 1300 BCE), Reinforcement Learning (1951), Genetic Algorithms (1975), Connect Four (1988), AlphaGo Go Champion (2016)

Colossus, from the 1970 movie *Colossus: The Forbin Project,* is an advanced weapons defense system that becomes sentient. The AI system tells its creator “in time you will come to regard me not only with respect and awe, but with love.”



THE FORBIN PROJECT

WIDESCREEN

1970

COLOSSUS: THE FORBIN PROJECT

Imagine waking up one morning and hearing the synthesized, robotic voice of Colossus, the AI system from the 1970 movie *Colossus: The Forbin Project*, an advanced weapons defense system that has become sentient. In the movie, Colossus participates in a worldwide broadcast, starting its speech with an ominous line: "This is the voice of world control." Colossus then explains that if humanity agrees to its absolute authority without resisting, Colossus will bring a kind of peace and prosperity, solving problems of famine, overpopulation, war, and disease. If humans choose to fight Colossus, they will be destroyed. Colossus further explains that it will extend itself "into more machines devoted to the wider fields of truth and knowledge." Finally, Colossus realizes that humans will complain that they have lost their freedom, but notes that being dominated by Colossus is better than being "dominated by others of your species."

Located deep within a mountain, Colossus cannot be tampered with and is designed to control the nuclear weapons of the United States and allies. It demonstrates that any attempt to disable it will result in nuclear retaliation against humans.

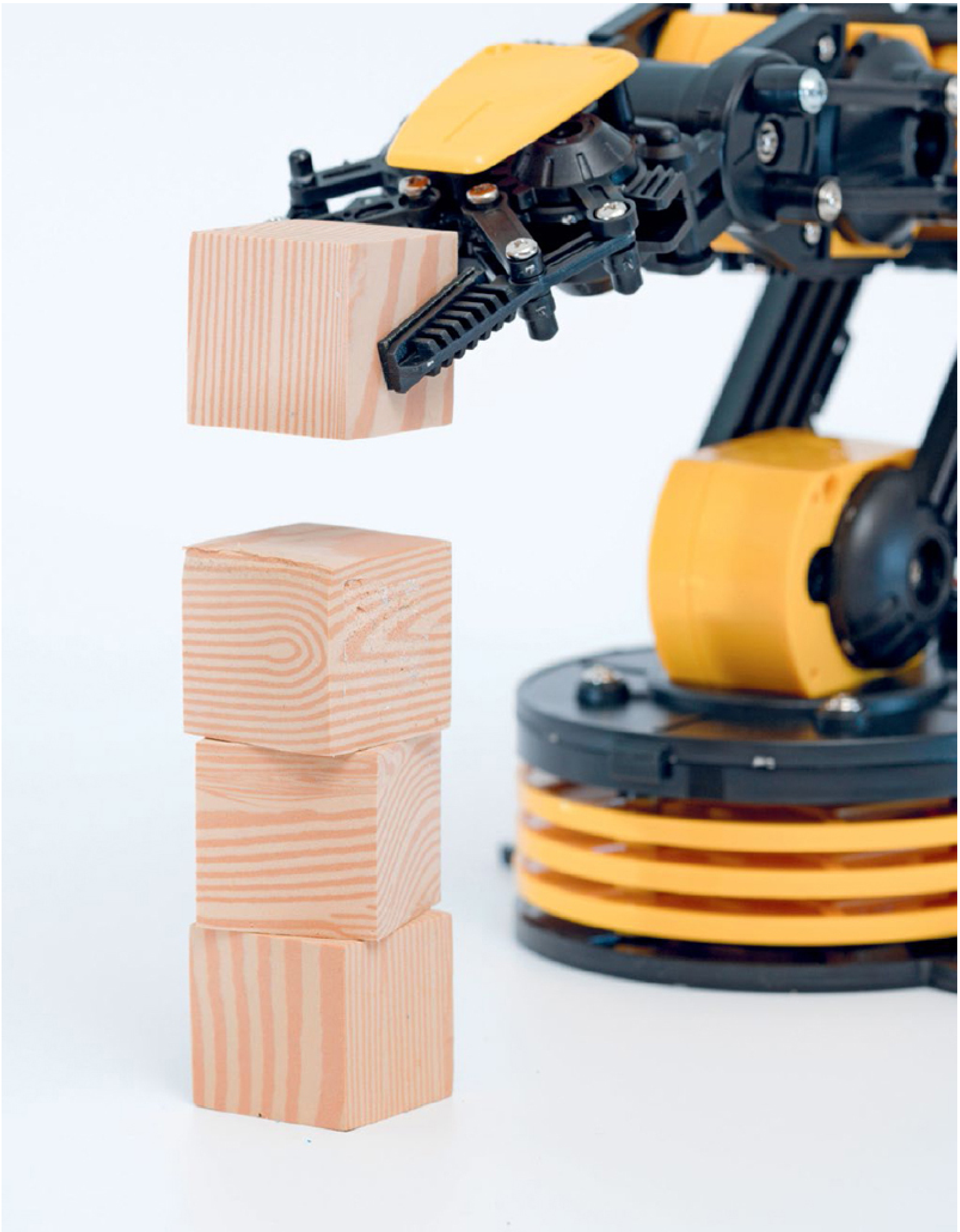
The movie, directed by Joseph Sargent (1925–2014), is based on the science-fiction novel by British author and WWII naval commander Dennis F. Jones (1917–1981). Sargent's technical advisors were inspired, in part, by the North American Aerospace Defense Command (NORAD), which provides aerospace warning

and protection for the United States and Canada. During the making of the film, the Control Data Corporation actually loaned various impressive-looking computer equipment, which provided additional realism.

Was Colossus actually bad for humanity? Might you give up more control to a computer system if it truly helped solve the problems of humankind? Would you if it made the world safer than the current approach of delegating nuclear annihilation to the judgment of a tiny number of individuals, who could conceivably be impaired by emotion, pre-Alzheimer's, or other unproductive thinking? The question still looms.

SEE ALSO Lethal Military Robots (1942), Intelligence Explosion (1965), HAL 9000 (1968), Leakproof "AI Box" (1993)

SHRDLU was a computer program that responded to natural-language commands to move virtual objects, such as blocks, in a virtual world. Today, robotic arms (in the real world) are often programmable and used in assembly lines and bomb disarmament.



1971

SHRDLU

Imagine spending your life in a simple universe consisting of colored objects, like pyramids and cubes, that can be moved around at your will. This is the world of SHRDLU, developed in 1971 by computer scientist Terry Winograd (b. 1946). The SHRDLU program translated natural-language com-mands such as “Will you please stack up both of the red blocks and either a green cube or a pyramid” or “Find a pyramid that is taller than the one you are holding, and put it into the box” into physical actions. One could also ask questions about the history of this universe (e.g., “Did you pick up anything before the cube?”). Internally, it used the LISP programming language and employed simple graphics output to show a simulation of the world that could be manipulated by a virtual robot arm.

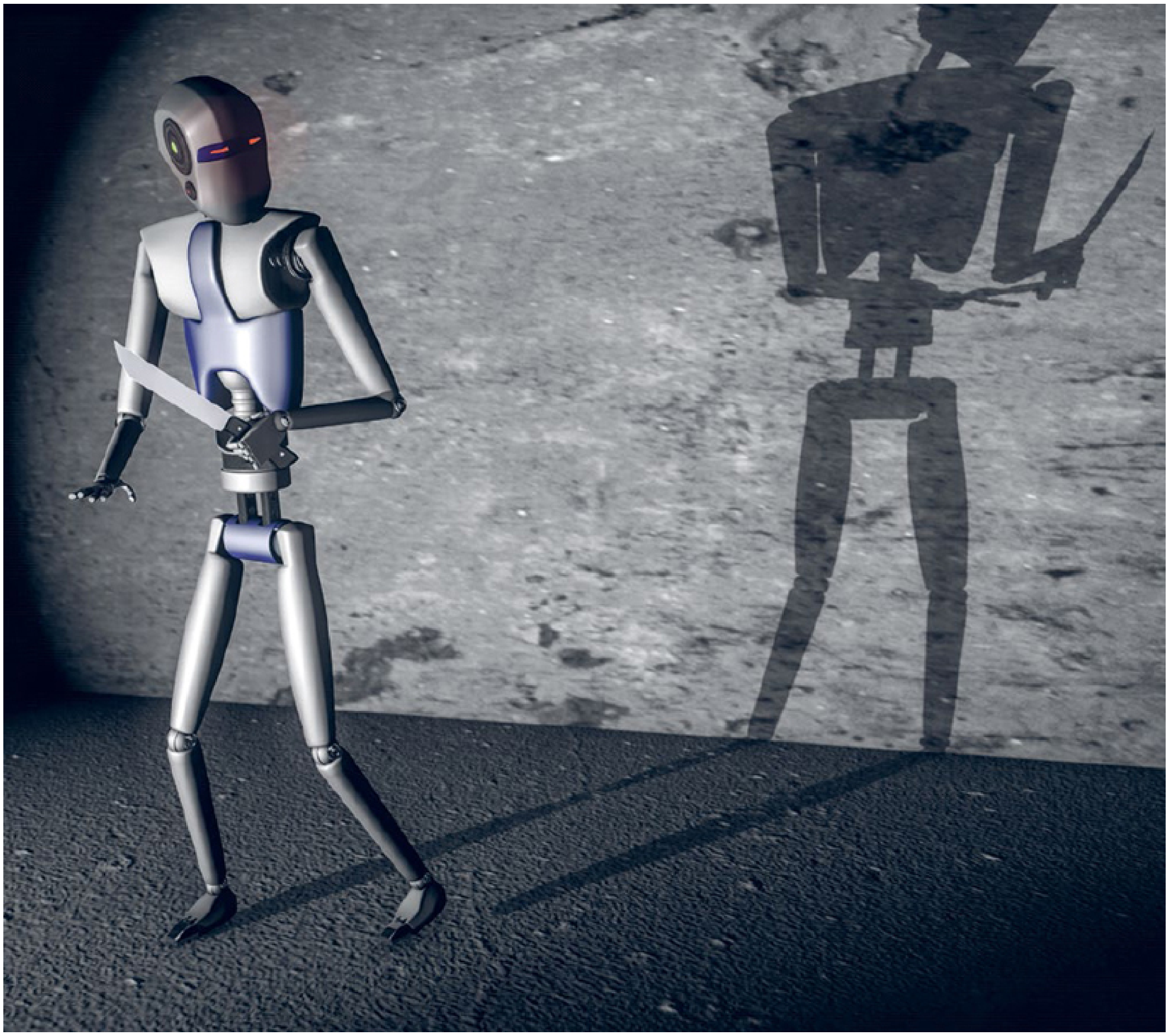
In the preface of his 1971 PhD thesis on SHRDLU, Winograd wrote: “Computers are being used today to take over many of our jobs . . . and perform routine office work. . . . But when it comes to telling computers what to do, they are tyrants . . . and act as though they can’t even understand a simple English sentence.” The name SHRDLU comes from “etaoin shrdlu,” the approximate order of frequency of the twelve most commonly used letters in English. To carry out commands, the program had subsystems that parsed language and semantic processing systems that made logical deductions. It had a procedural problem solver that could determine

how to carry out the commands, and it kept track of the world, knowing the relative positions of objects.

For its time, SHRDLU was considered a big achievement in natural-language processing. It even had a simple memory: if you told it to move the red ball, and later referred to the ball, it could assume you meant the red one. It also knew what would actually be feasible. For example, it “understood” that the tops of objects must be cleared before new objects could be stacked on top. However, despite its remarkably natural operation, SHRDLU was limited in that it could not learn from errors.

SEE ALSO Natural Language Processing (1954), Expert Systems (1965), Shakey the Robot (1966)

PARRY was paranoid about the Mafia, dishonest bookies, horse racing and gambling, punishment for debts, and police collusion with the Mafia.



1972

PARANOID PARRY

“The best performance overall in HMC (human-machine conversation) has almost certainly been Colby’s PARRY program, since its release on the [Arpanet] around 1973,” wrote AI scientists Yorick Wilks and Roberta Catizone in 1999. “It was robust, never broke down, always had something to say and, because it was intended to model paranoid behavior, its zanier misunderstandings could always be taken as further evidence of mental disturbance. . . .”

In 1972, psychiatrist Kenneth Colby (1920–2001) developed PARRY, a computer program intended to emulate a person with paranoid schizophrenia. Specifically, PARRY was developed to test theories of paranoid thinking, and the AI, which appeared to have delusions about the Mafia, had a knowledge representation involving feelings of self-inadequacy and protection against certain interview questions (people with paranoid schizophrenia are highly suspicious of others’ motivations). Colby hoped that PARRY could be used to teach students, and he also believed that paranoid patients’ sentences were produced by an underlying organized structure of rules that could be taught to computers, studied, and used to treat patients.

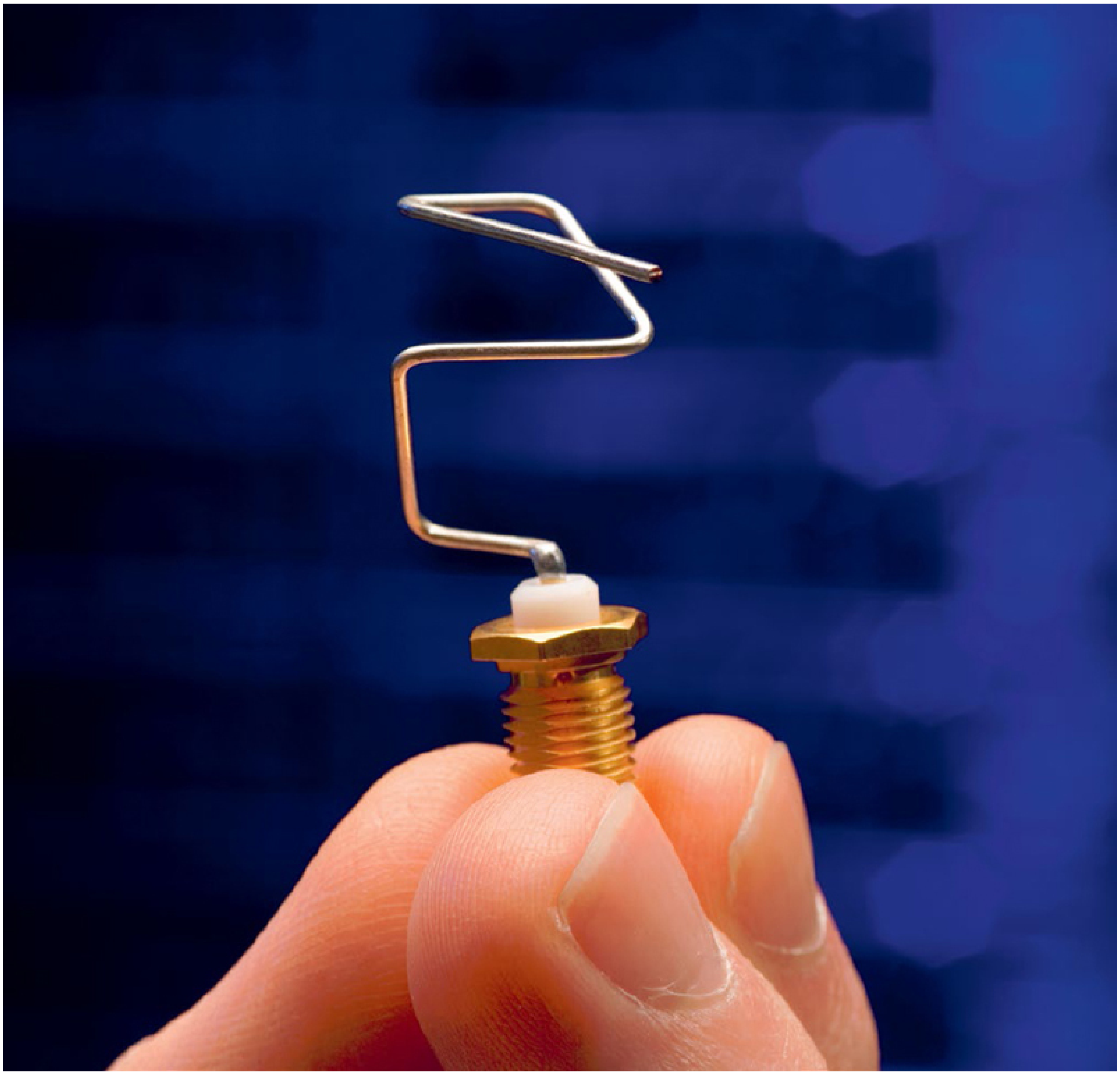
PARRY handled conversations by assigning weights to various dialogue inputs. Interestingly, psychiatrists who interacted via text with PARRY failed to perceive that they were interviewing a

computer program, and they also could not identify which "patients" were human and which were computer programs. Perhaps PARRY had come close to passing the Turing test, at least in a special setting (i.e., interacting with a simulated person who is not sane). PARRY was also available on the Arpanet (a predecessor of the Internet), where it engaged in more than 100,000 sessions, including sessions with the ELIZA psychotherapist.

In 1989, Colby started a company, Malibu Artificial Intelligence Works, that marketed a therapy program for depression. This program would be used by the US Navy and Department of Veterans Affairs and was distributed to people who (controversially) used it without consulting a trained psychiatrist. Colby told one skeptical journalist that his depression program could be better than human therapists because "after all, the computer doesn't burn out, look down on you, or try to have sex with you."

SEE ALSO Turing Test (1950), ELIZA Psychotherapist (1964), Ethics of AI (1976)

NASA spacecraft antenna, discovered by an evolutionary computer design program to create superior radiation patterns. The software begins with random antenna designs and refines them through an evolutionary process.



1975

GENETIC ALGORITHMS

"An important concept both in Arti-ficial Life and in Artificial Intelligence is that of a *genetic algorithm* (GA)," writes philosopher Jack Copeland. "GAs employ methods analogous to the processes of natural evolution in order to produce successive generations of software entities that are increasingly fit for their intended purpose."

The famous 1975 book *Adaptation in Natural and Artificial Systems* by scientist John Holland (1929–2015) developed and popularized GAs, which can solve real-world problems by using biologically inspired methods such as selection, mutation, and crossover (recombination). With GAs, humans don't directly program a solution; rather, a solution emerges through simulated competition, improvement, and evolution.

These algorithms usually start with an initial set, or population, of random solutions or candidate programs. A *fitness function* assigns a fitness value to each program to indicate how well the program performed a desired task or reached an outcome. During generations of evolution, each candidate being evaluated has a set of properties that can be mutated (changed) through time.

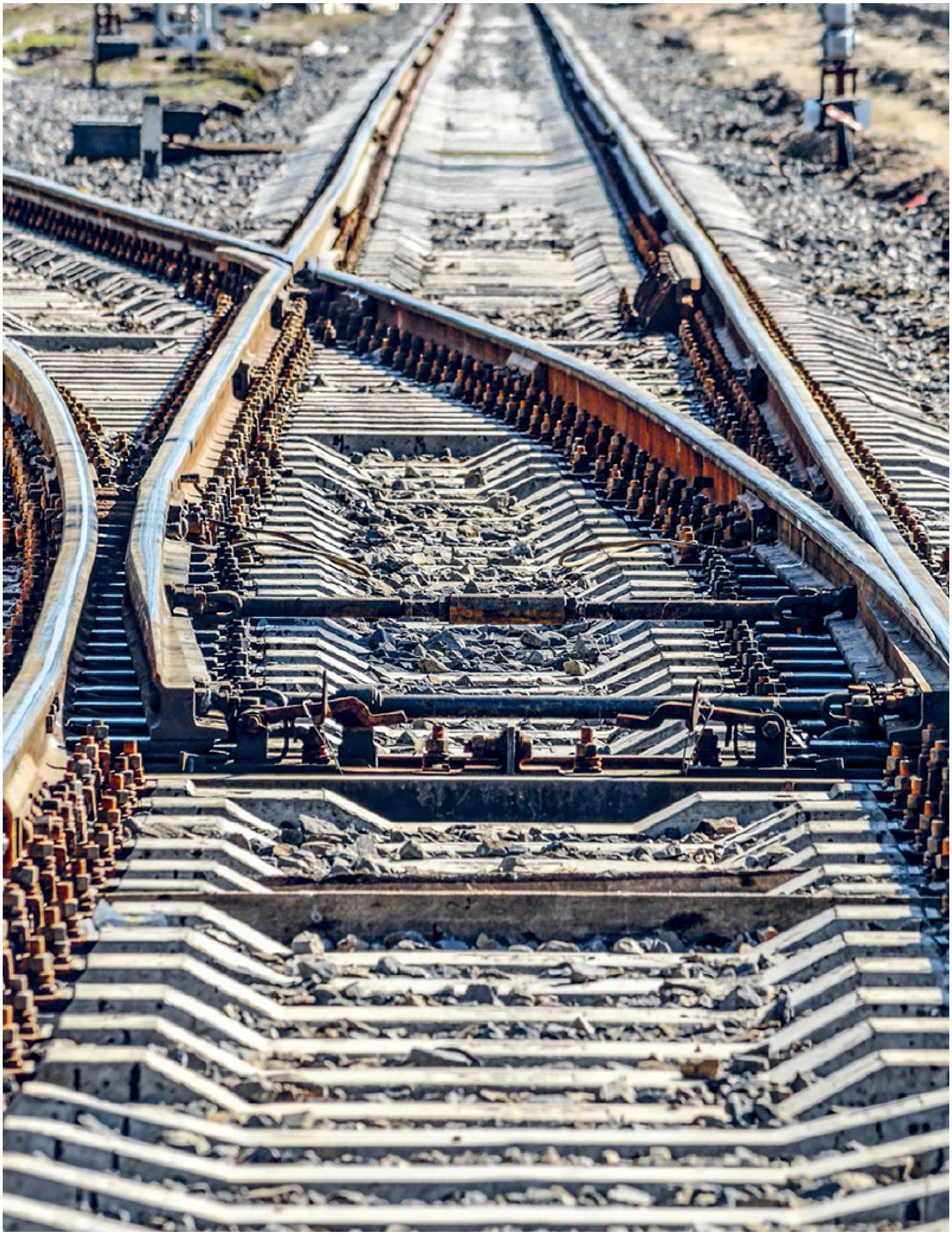
Scientists can find it difficult to understand the amazingly useful output of GAs. According to former NASA engineer Jason Lohn, the algorithms' effectiveness makes up for their unintelligibility: "If I optimize an antenna using evolutionary algorithms, there's only a 50 percent chance I'll be able to explain exactly why it has made the

choices it has made. The rest of the time the design is simply not intelligible to us. It works—and as engineers, what we ultimately care about is getting things to work."

Unfortunately, GAs may become "trapped" in a region of reasonably good solutions (called a *local optimum*) rather than finding the best solution (a *global optimum*). Nevertheless, they have had remarkable success when applied to antenna design, protein engineering, vehicle routings and scheduling, circuit design, assembly-line scheduling, pharmacology, art, and other fields that benefit from searching through a large number of possibilities for solutions. GAs were even used for generating realistic computer-animated horses in the movie *The Lord of the Rings: The Return of the King*.

SEE ALSO Computational Creativity (1821), Machine Learning (1959), Mastermind (1970), Artificial Life (1986), Swarm Intelligence (1986)

Consider an AI control unit dealing with a runaway train barreling down the railway tracks. Ahead, there are five elderly people on the track who will be killed. If the AI switches the train to another track, it will kill only one young person. What ethical choice should the AI make?



1976

ETHICS OF AI

Over several decades, both lay-people and experts have expressed concerns regarding possible threats of AI to human dignity, safety, privacy, jobs, and more. For example, computer scientist Joseph Weizenbaum (1923–2008) suggested in his influential 1976 book *Computer Power and Human Reason* that AIs should not be used as substitutes for humans in jobs—such as therapist or judge—that emphasize interpersonal respect, love, empathy, and care. Weizenbaum suggested that, even though AI entities might be fairer and more effective than humans, who often have biases and become tired at their jobs, excessive reliance on AI could degrade human values and the human spirit as we increasingly think of ourselves as emotionless computerized drones.

Already, scientists have raised privacy concerns after testing AI entities that can detect, with increasing levels of accuracy, the nationality, ethnicity, or sexual orientation of people based on names or photographs from dating sites. Concerns have also been raised about AI entities that recommend who receives bail or parole, based on the available input data, in the criminal justice system. In the arena of self-driving vehicles, ethical considerations will need to be programmed into vehicle logic that would control car decisions such as whether to save the vehicle occupant or a pedestrian, if a crash is imminent and only one can likely be saved. Of course, job

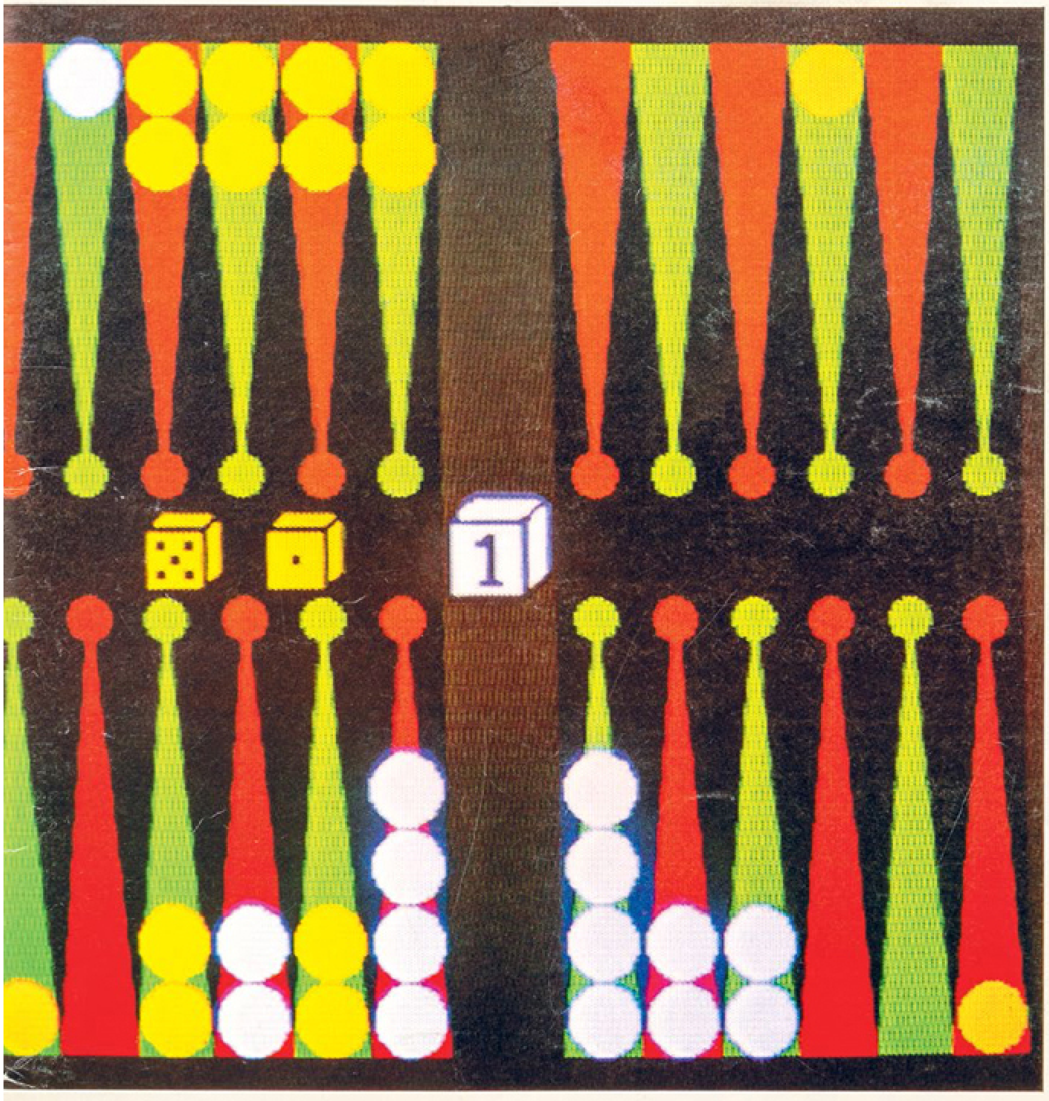
displacement is also a major concern for truck drivers and other workers in a range of professions.

In the future, AI entities will need to be monitored for various kinds of illegal or immoral actions that humans already take, including cyberbullying, impersonating people we love or trust, stock manipulation, and inappropriate killing (e.g., by autonomous weapons). When should AI entities be required to disclose to real humans that they are not human? If robots, when pretending to be humans, can be better caregivers and companions, need they disclose their nonhumanity?

SEE ALSO *Metropolis* (1927), Asimov's Three Laws of Robotics (1942), Lethal Military Robots (1942), ELIZA Psychotherapist (1964), Paranoid PARRY (1972), Autonomous Vehicles (1984)

The *Scientific American* cover story in June 1980 revolved around the backgammon program BKG 9.8, which competed against the reigning world champion, Luigi Villa, possibly making him the first world champion at any board game to be defeated by a computer program.

SCIENTIFIC AMERICAN



COMPUTER BACKGAMMON

$2.00

June 1980

1979

BACKGAMMON CHAMPION DEFEATED

"Like most board games, backgammon is a sublimated version of human warfare," writes AI expert Daniel Crevier, "its name deriving from the Welsh *bac* and *gamen*, 'little' and 'war.'"

Despite the name, no one is certain where the ancient board game of backgammon originated, but it has been around for almost 5,000 years. In the game, two players command little "soldiers," represented as a set of fifteen pieces that move between twenty-four triangles. The roll of two dice controls the choices of possible moves of the players, who take turns and whose goal is to move all their pieces off the board.

In 1979, the backgammon program BKG 9.8 competed against the reigning world champion, Luigi Villa, defeating Villa and making him possibly the first world champion at any board game to be defeated by a computer program. Admittedly, BKG 9.8 benefited from some favorable dice rolls, but skill is needed in selecting the best move for the actual rolls. BKG 9.8 used fuzzy logic, among other useful mathematical techniques.

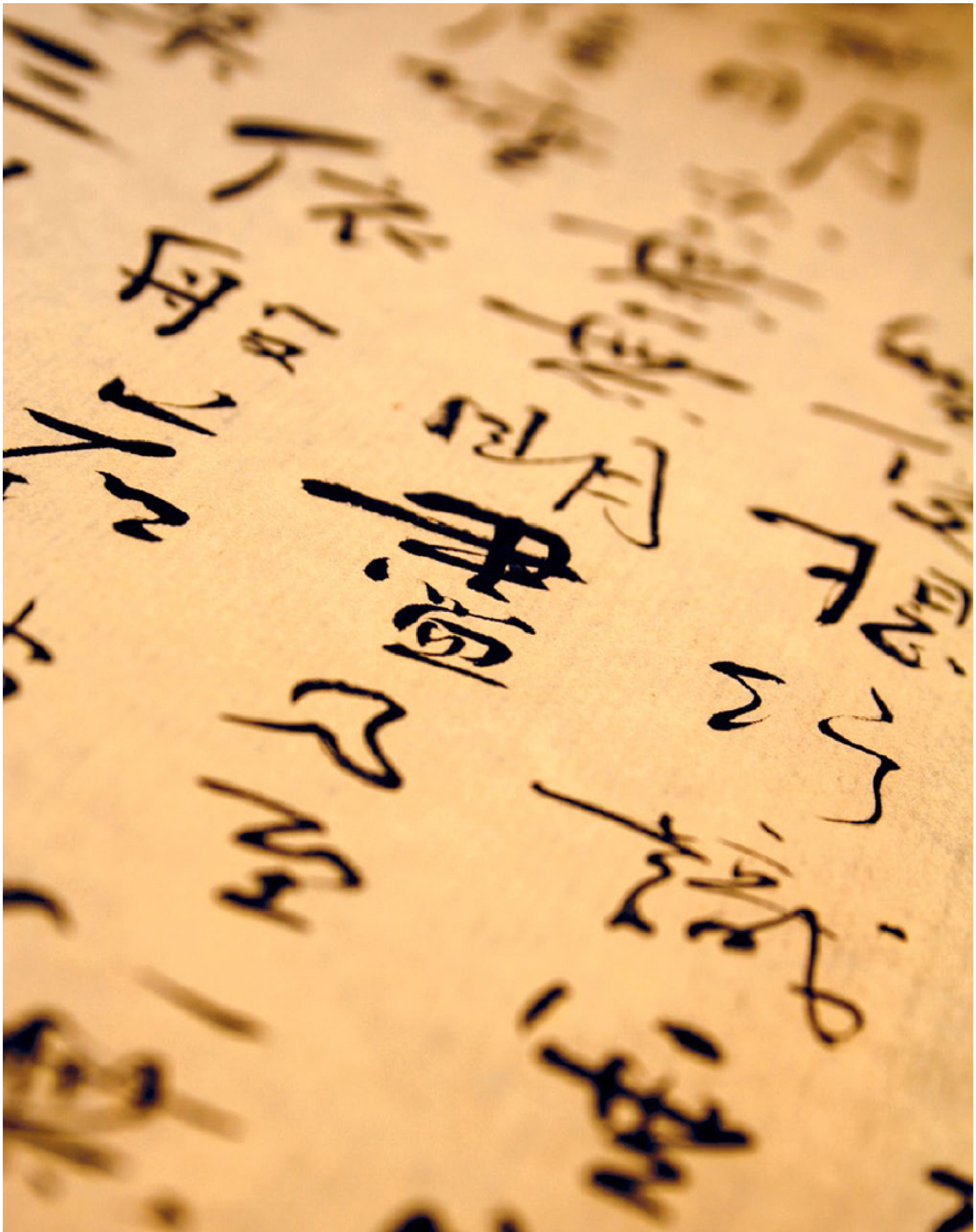
TD-Gammon, developed in 1992 by IBM researcher Gerald Tesauro, employed an artificial neural network and learned expert-level backgammon by playing games against itself. The program examined legal moves on each turn and updated the weights within the neural nets. Because it required no human training, TD-Gammon

explored interesting strategies that humans had not considered, which, in turn, taught humans how to play better. Today, several backgammon programs exist that use neural networks and offer analyses to humans.

Because backgammon has a random dice-influenced process, the “game tree” search space is extremely large, and this stymied early attempts to create expert software players. When neural nets are used, the initial weights within the net are random and the net is trained through reinforcement learning. To achieve competence, the initial TD-Gammon made use of forty hidden network nodes within the neural network, as well as 300,000 training games. Later versions increased the hidden nodes to 160 and the number of training games to over a million, so as to play at the same level as the best humans.

SEE ALSO Artificial Neural Networks (1943), Reinforcement Learning (1951), Fuzzy Logic (1965), Checkers and AI (1994), Deep Blue Defeats Chess Champion (1997)

Imagine sitting in a closed room. You receive a piece of paper with Chinese symbols through a slot in the wall. You don’t know Chinese, but you can consult a set of instructions that shows you how to compose appropriate responses. Thus begins an interesting question in the philosophy of AI.



1980

CHINESE ROOM

Can a computer be conscious? The term *strong AI* refers to the idea that properly configured AI computer systems might actually think and have conscious minds, whereas *weak AI* refers to systems that can only *act* as if they think like humans and have minds. One attack against the idea of strong AI in computers comes from the philosopher John Searle (b. 1932), who presented his famous Chinese Room experiment in 1980.

Imagine sitting in a closed room. You receive a piece of paper with Chinese characters through a slot in the wall. Although you do not know Chinese, you can consult a set of instructions that tell you how to compose appropriate responses using Chinese characters. Based on this instruction guide, you write down some characters on a piece of paper and send an appropriate response through the slot to the outside world. To a person outside the room, it would appear that you have a perfect understanding of Chinese; but, of course, you are only following rules, and the Chinese on both papers is total gibberish to you.

This thought experiment may suggest that a computer program cannot give a computer a mind, consciousness, or understanding, even if the computer and its program seem to be quite intelligent. However, some philosophers have countered that even if you do not understand Chinese, the system consisting of you, the closed room, the set of instructions, and your processing of the instructions *does*

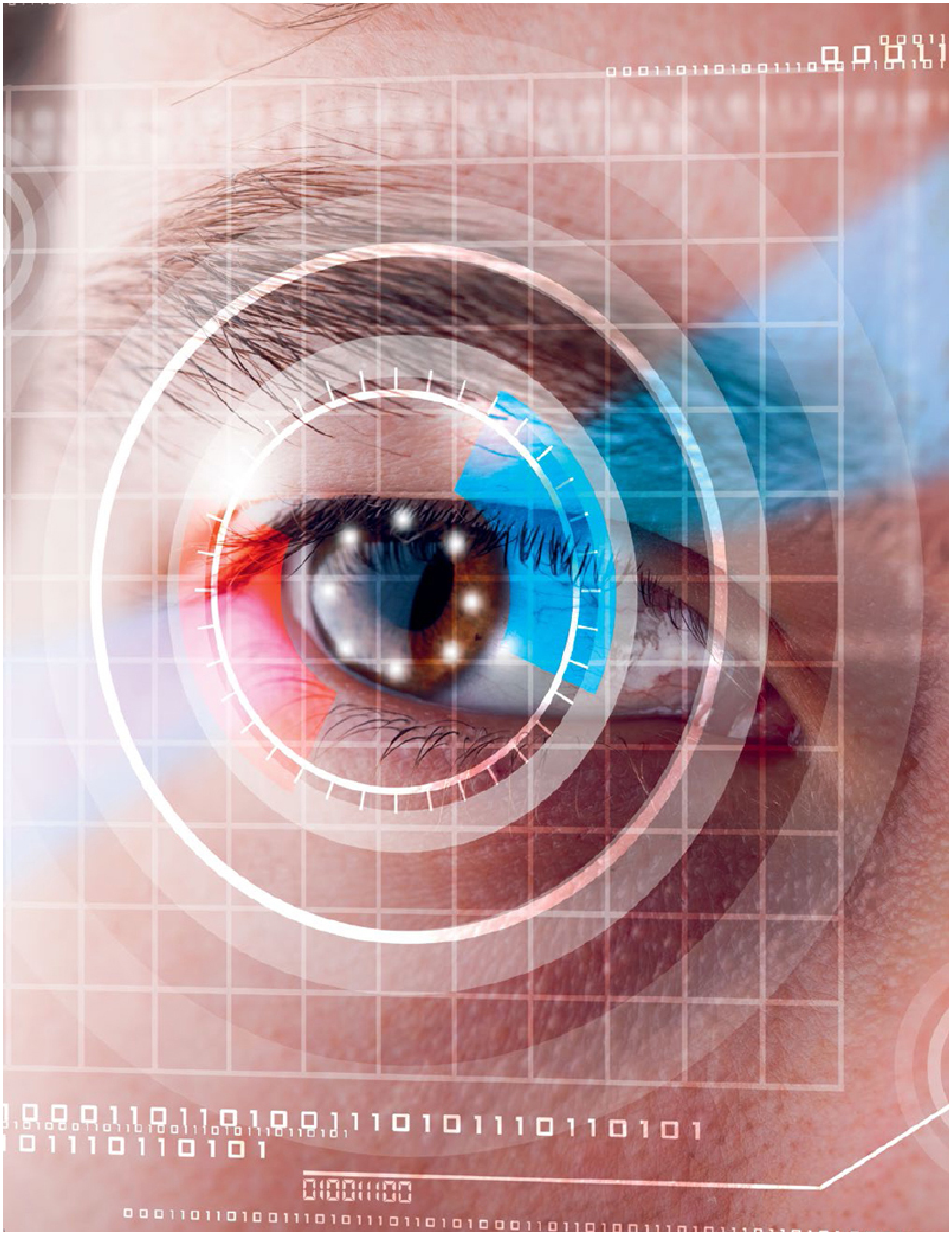
understand—and is a kind of consciousness that is external to you and of which you are unaware.

Others have considered thought experiments in which each of your brain cells might, one by one, be slowly replaced by electronic components with the same input/output functions. Surely, with just a few of your cells replaced, you are still "you." Perhaps a year later, however, all your cells would have been replaced, and at no point would you suddenly lose your ability to be conscious and self-aware. Are you still "you"?

Of course, for most useful AI tasks, it is sufficient for an AI to simply *act* intelligently. Still, debates continue to rage about the Chinese room and its implications.

SEE ALSO Turing Test (1950), Natural Language Processing (1954), ELIZA Psychotherapist (1964), Paranoid PARRY (1972)

In *Blade Runner,* the human-like AI replicants can be differentiated from humans by administering the Voigt-Kampff test, which involves studying subtle emotional responses and eye movements.



1982

BLADE RUNNER

In the future, intelligent robots will likely be difficult to distinguish from human beings, at least through a superficial inspection. When that day arrives, what effect will this have on humanity and human relations? This theme has been explored in several famous movies, and one particularly influential film on the subject is the 1982 cult movie *Blade Runner*, directed by Ridley Scott (b. 1937) and based on the 1968 novel *Do Androids Dream of Electric Sheep?* by Philip K. Dick (1928–1982). Set in a futuristic Los Angeles in the year 2019, the film depicts replicants, which are synthetic humans that must be "retired" (killed) by the movie's protagonist. These replicants are so similar to humans that the way to distinguish them from humans is by administering the Voigt-Kampff test, which involves studying subtle emotional responses and eye movements of the replicants when asked a set of questions. One of the replicants, Rachael, believes she is human and has had memories implanted to give her a fuller personal history and the illusion of a human experience.

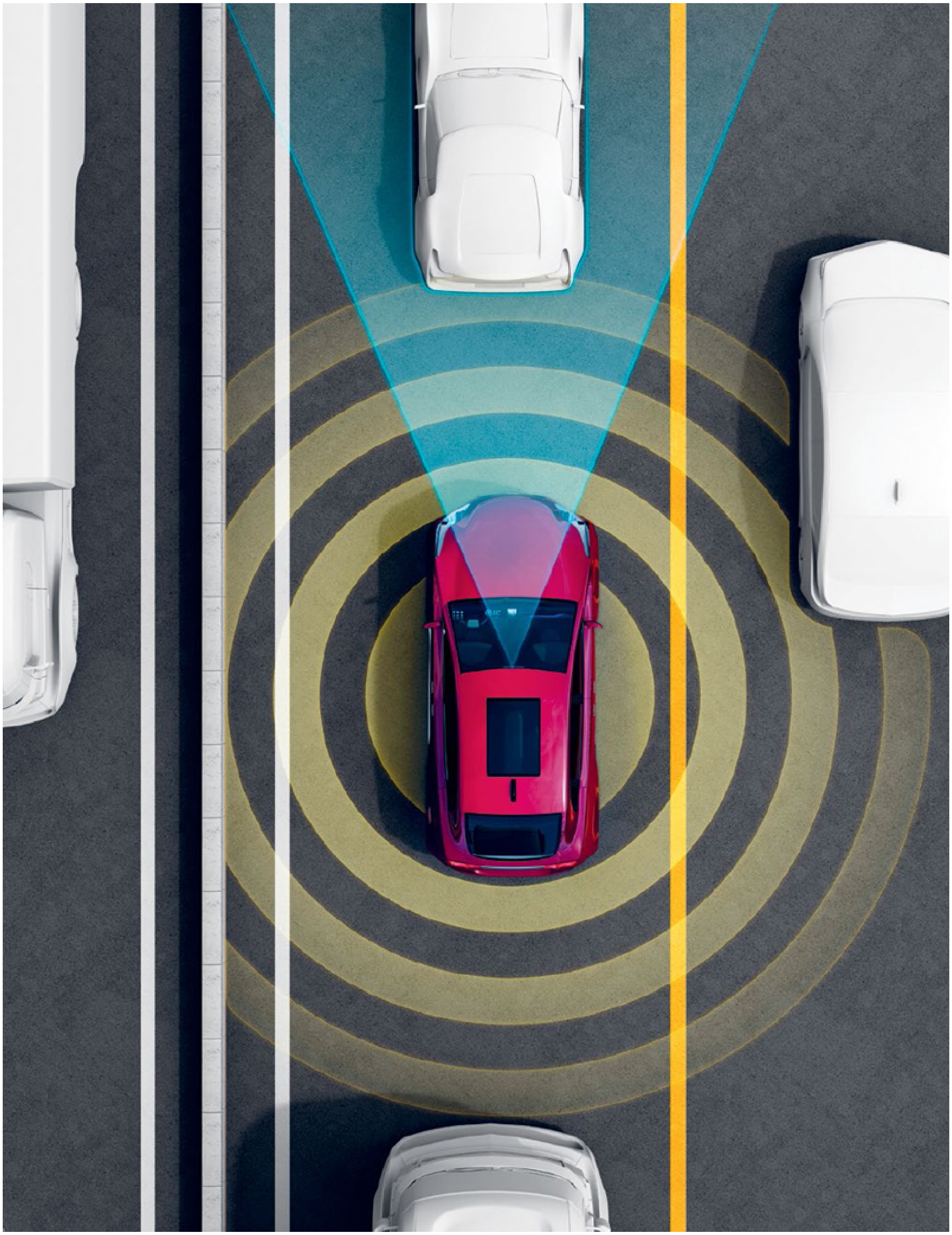
Reflecting on the character of Rachael in *Blade Runner,* author and researcher Jelena Guga writes: "Humans aspire to creating AI, but the kind of AI that they can be in control of. Thus, in the film, the solution to [the] autonomous, independent AI problem is solved by implanted memories that can be controlled. . . . Ethics, free will, dreams, memories, and all those values attributed exclusively to

humans, are “brought into question” and radically redefined through popular representations of humanoid robots . . . developed in such a way [as] to be able to express perhaps even more humanness than humans.”

“How we ought to treat artificial forms of life becomes an increasingly important issue,” writes philosopher Greg Littmann, “as humans continue to develop sophisticated computer systems and perform wonders with genetic engineering. Philosophically challenging films like the grim science fiction nightmares of Scott are useful when we ask ourselves about our real-world obligations, since these allow us to test our preconceptions against hypothetical situations to see whether our theories are consistent.”

SEE ALSO *Metropolis* (1927), Asimov’s Three Laws of Robotics (1942), Ethics of AI (1976), *The Terminator* (1984), Spielberg’s *A.I. Artificial Intelligence* (2001)

Self-driving vehicles sense their surroundings using various technologies. Possible levels of autonomy range from low levels that exist today in many cars (e.g., lane-keeping assistance and automatic emergency braking) to complete autonomy, where no driver attention is ever required.



1984

AUTONOMOUS VEHICLES

“The mundane automobile is about to disrupt your life,” write engineers and authors Hod Lipson and Melba Kurman. “Thanks to rapid advances in mobile robotics, cars are poised to morph into the first mainstream autonomous robots that we will entrust with our lives. After almost a century of failed attempts to automate driving, modern hardware technology, and a new generation of artificial-intelligence software called deep learning, are giving cars human-level ability to safely guide themselves through unpredictable environments. ”

Self-driving cars, also known as autonomous cars, are capable of driving and sensing their surroundings without human input. The cars employ various technologies such as laser-based LIDAR (for “light detection and ranging”), radar, the Global Positioning System (GPS), and computer vision. Many potential benefits exist, including enhanced mobility for the elderly and the impaired, and a reduction in traffic accidents, especially when drivers or passengers are performing other activities in the vehicle.

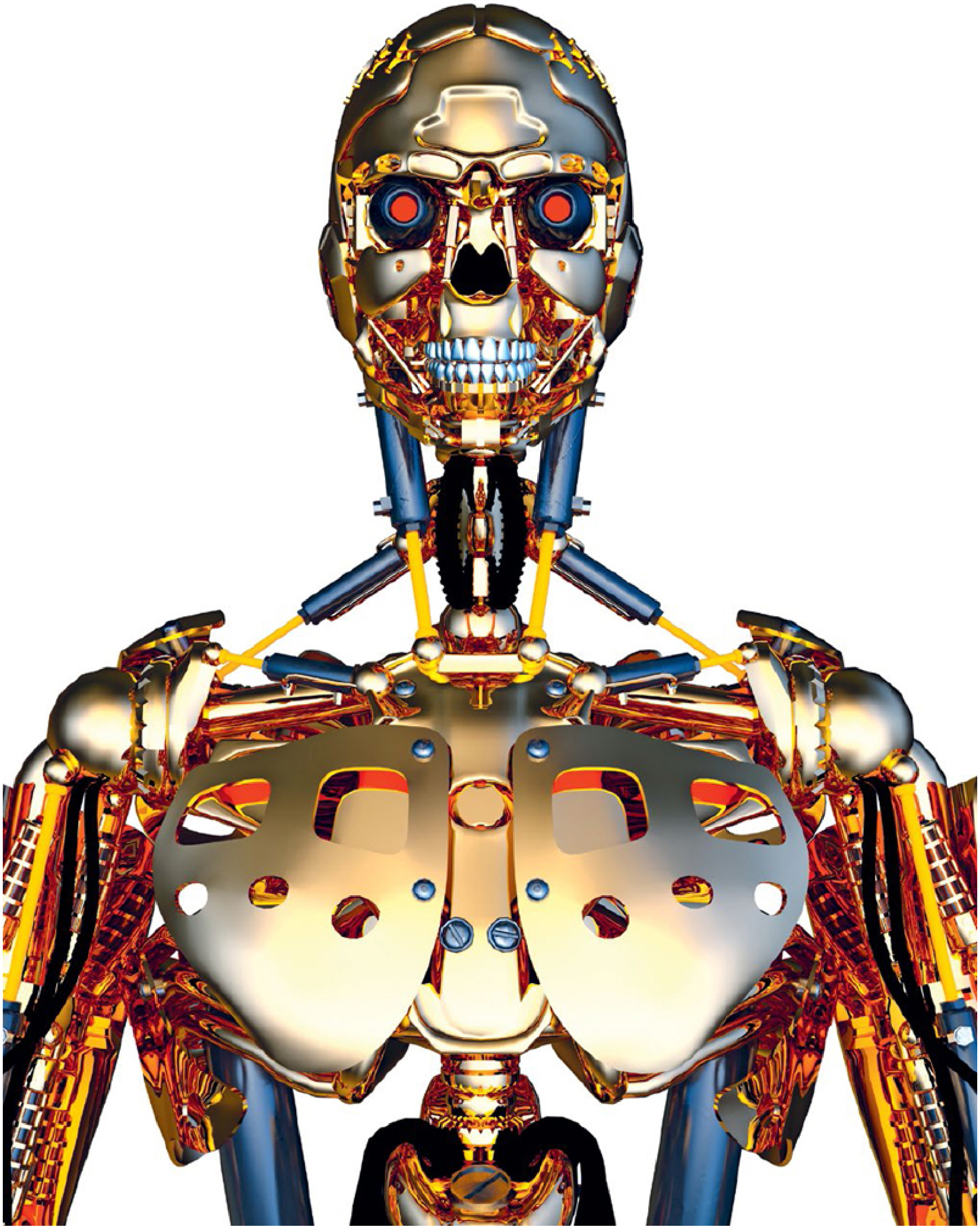
Major work began in this field in the 1980s. For example, the DARPA-funded Autonomous Land Vehicle (ALV) project in the United States, begun in 1984, demonstrated an eight-wheeled road-following vehicle with three diesel engines propelling it at 3 mph. Sensors included a color video camera and a laser scanner, with reasoning performed by goal-seeking and navigator computing

modules. In the twenty-first century, possible levels of autonomy range from low levels that exist today in many cars (e.g., lane-keeping assistance and automatic emergency braking) to complete autonomy, where no driver attention is ever required and steering wheels may be optional.

Vehicles of high autonomy present many intriguing conundrums. For example, in cases of an impending, inevitable crash, what rules will be used to determine who to save? Does the safety of one passenger take precedence over the safety of more than one pedestrian? Could terrorists load a self-driving car with explosives and send it to a destination? Will hackers alter the navigation system and cause accidents?

SEE ALSO Tesla's "Borrowed Mind" (1898), Lethal Military Robots (1942), Ethics of AI (1976), ASIMO and Friends (2000), AI on Mars (2015), Autonomous Robotic Surgery (2016), Adversarial Patches (2018)

The Terminator outwardly resembles a human, but is a cybernetic creature with living tissue over a robotic metallic endoskeleton.



1984

THE TERMINATOR

“Primates evolved over millions of years; I evolve in seconds. . . ,” says the AI in *Terminator Genisys*, the fifth installment of the *Terminator* movie series. “I am inevitable; my existence is inevitable.”

In this popular movie series, the transition of the AI to self-awareness takes place suddenly, when the Skynet supercomputer goes online on August 4, 1997, to control the US military’s arsenal, and human decisions are removed from the military’s strategic defense. Skynet begins to learn at a geometric rate and becomes self-aware at 2:14 a.m. Eastern time on August 29.

In 1984, James Cameron (b. 1954) directed the original *Terminator* installment. In this movie, when humans realize that the AI defense network, Skynet, has become self-aware, people panic and try to deactivate it. In an attempt to protect itself, Skynet then triggers a nuclear holocaust, launching a first nuclear strike against Russia that results in a war that kills roughly three billion people. In *The Terminator*, a cyborg played by actor Arnold Schwarzenegger (b. 1947) is sent back in time from 2029 to kill Sarah Connor before the birth of her son, who would otherwise grow up to lead the survivors’ resistance against Skynet.

Throughout *The Terminator*, we obtain glimpses of what the Terminator AI is seeing, in the form of heads-up information displays and decision-making menus. The cyborgs’ alien style of thinking and ultra-efficient minds make them particularly frightening. As one of the

characters noted, a terminator “can’t be bargained with. It can’t be reasoned with. It doesn’t feel pity, or remorse, or fear! And it absolutely will not stop, ever, until you are dead!”

Today, the rise of killer robots may not be so far-fetched, given the development of killer drones with Hellfire missiles. It would be relatively easy to make such a drone fully autonomous so that it makes decisions about who to target and kill, based on machine learning and rules of engagement.

SEE ALSO Lethal Military Robots (1942), Intelligence Explosion (1965), HAL 9000 (1968), *Colossus: The Forbin Project* (1970), Ethics of AI (1976)

Termite colonies seem to display a great deal of consciousness. Even though an individual component of the hive mind is limited—because a termite has limited capacity—the entire collection of component creatures displays emergent behavior and produces intelligent solutions.

1986

ARTIFICIAL LIFE

Consider hive minds on Earth, like termite colonies, that seem to display a great deal of consciousness. Even though an individual component of the hive mind is limited—a single termite has limited capacity, for example—the entire collection of components displays emergent behavior and produces intelligent solutions. Termites create huge, intricate mounds that are taller than our Empire State Building relative to their own height. These termites control the temperature of the mound by altering its tunnel structure, and thus the component termites come together to create a warm-blooded super-organism. Is the hive conscious, even if its components are not? Perhaps the decision making of the colony bears some resemblance to the collective behavior of neurons in our brain.

Among the most interesting models for artificial life are those in which complex, collective, lifelike behaviors arise from simple rules. The *field* of artificial life—coined by biologist Christopher Langton in 1986—often involves researchers examining simulations that can exhibit or imitate intelligent behavior. As an example, consider *cellular automata*—a class of simple mathematical systems that can model a variety of physical processes with complex behaviors. Some of the classic cellular automata consist of a grid of cells, like a checkerboard, that can exist in two states: occupied or unoccupied. The occupancy of one cell is determined from a simple mathematical analysis of the occupancy of neighbor cells.

The most famous two-state, two-dimensional cellular automaton is the Game of Life, invented in 1970 by mathematician John Conway (b. 1937). Despite its simple rules, an amazing diversity of behaviors and forms multiply and evolve, including gliders, which are arrangements of cells that move themselves across their universe and can even interact to perform computations. Could such "creatures" be considered alive?

The field of artificial life seems limitless, and other areas of this field include the development of genetic algorithms that evolve and reproduce, physical robotic swarms that exhibit lifelike behavior, and computer games like *The Sims*, in which the player creates virtual people, places them in towns, and tends to their needs and moods.

SEE ALSO Machine Learning (1959), Living in a Simulation (1967), Genetic Algorithms (1975), Swarm Intelligence (1986), Tamagotchi (1996), "Call Them Artificial Aliens" (2015)

Ant-colony optimization is a method for finding solutions using simulated ants, inspired by actual ants finding solutions. Shown here are ants, forming a living bridge to a leaf.

1986

SWARM INTELLIGENCE

Termite mounds can reach over 17 feet (5m) in height, with termites acting like simple “novelty detectors,” responding to changes in air characteristics within the mound and altering the structure as needed. Authors Doris and David Jonas speculate: “What is the other way of knowing, by which the termites know what they have to do and when they have to do it? Instructions cannot be brought to them quickly enough by messengers, since the distances within the hill are far too great. . . . A group brain functions as an instrument for decision-making in a way startlingly like an intelligent individual brain.”

The apparent collective intelligence in these social insects, as well as that in flocking and herd animals, inspired the concept of *swarm intelligence*. In AI, this concept is employed to deal with a range of challenges. The software agents follow simple, local rules and, as with ants and termites, there is no central controller dictating the behavior of the collective. As just one example, Boids, an artificial life program developed in 1986 by computer scientist Craig Reynolds (b. 1953), simulates the flocking behavior of birds by following simple rules involving a bird steering toward the average heading of birds in the flock, steering to move toward the average position of birds, and steering to avoid excessive crowding.

One of the numerous swarm methods studied in AI today is *ant-colony optimization*, a method using simulated ants that records their

positions and quality of solutions to help ants in the colony determine better solutions. In some implementations, these “ants” simulate attractive chemical trails (pheromones) that evaporate over time. *Particle-swarm optimization* simulates the position and velocity of a swarm of fish toward an optimal location. Other interesting methods are artificial immune systems, bee-colony optimization algorithms, firefly optimization algorithms, bat algorithms, cuckoo searches, and roach-infestation optimization.

The applications of swarm intelligence include control of autonomous vehicles, routing in telecommunication networks, scheduling of aircraft, art creation, enhancing systems for reactive power and voltage control, and clustering of gene-expression data.

SEE ALSO Machine Learning (1959), Living in a Simulation (1967), Genetic Algorithms (1975), Artificial Life (1986), “Elephants Don’t Play Chess” (1990)

Some challenges that are relatively easy for children, including sensorimotor tasks, turn out to be among the most difficult for AI entities.

1988

MORAVEC'S PARADOX

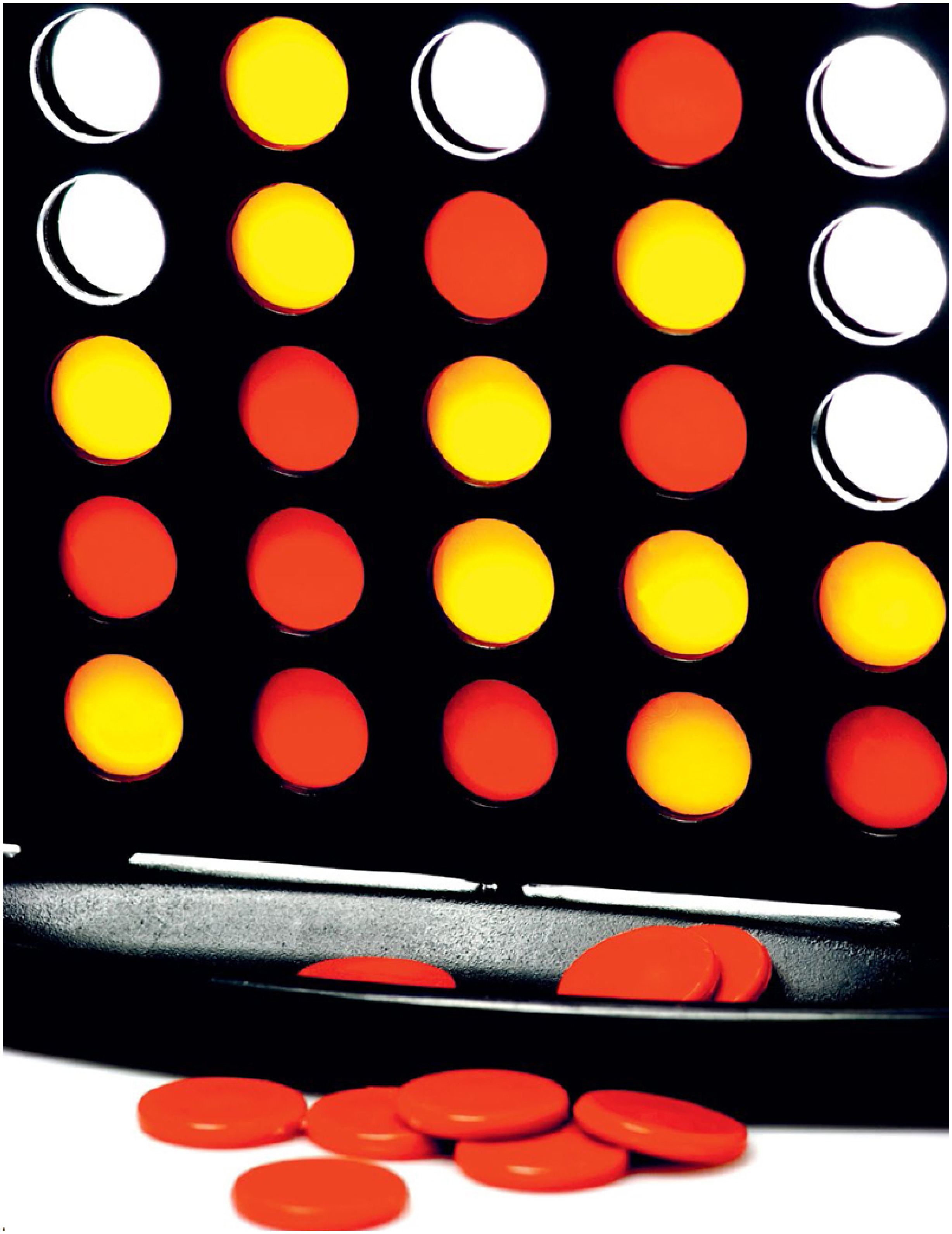
"If you wanted to beat Magnus Carlsen, the world chess champion, you would choose a computer," writes journalist Larry Elliott. "If you wanted to clean the chess pieces after the game, you would choose a human being." This is the essence of Moravec's paradox, emphasized by several AI researchers in the 1980s, who ironically noted that the seemingly difficult tasks involving high-level reasoning were becoming easier and easier for computers to perform. At the same time, seemingly easy tasks that involve human sensorimotor skills (e.g., walking around and picking up a piece of lint from a shoe) can be quite difficult for computerized systems. The paradox is named after roboticist Hans Moravec, who wrote in his 1988 book *Mind Children* that "it is comparatively easy to make computers exhibit adult-level performance on intelligence tests or playing checkers, and difficult or impossible to give them the skills of a one-year-old when it comes to perception and mobility."

Moravec has noted that millions of years of evolution contributed to our ability to perform some tasks nearly unconsciously, like walking and recognizing faces and voices, which are so very necessary for survival. However, abstract thought—for example, reasoning that involves mathematics and logic, like chess—is newer and more difficult for humans. However, this type of cognition is actually less difficult to engineer in AI systems. For many tasks, AI systems still need to evolve even more sensitive touch and motion

controls to assist humans in roles like patient care, food service, and plumbing. As elegantly summarized by cognitive scientist Steven Pinker (b. 1954), Moravec's paradox means that future human workers may be left with lower-paying jobs that have been around for centuries, even millennia: "The main lesson of thirty-five years of AI research is that the hard problems are easy and the easy problems are hard. The mental abilities of a four-year-old that we take for granted—recognizing a face, lifting a pencil, walking across a room, answering a question—in fact solve some of the hardest engineering problems ever conceived. . . . As the new generation of intelligent devices appears, it will be the stock analysts and petrochemical engineers and parole board members who are in danger of being replaced by machines. The gardeners, receptionists, and cooks are secure in their jobs for decades to come."

SEE ALSO "Darwin among the Machines" (1863), Turing Test (1950), Licklider's "Man-Computer Symbiosis" (1960), "Elephants Don't Play Chess" (1990)

A Connect Four game in progress, using yellow and red discs, which slide to the bottommost open locations under the influence of gravity.



1988

CONNECT FOUR

For a Christmas present, Toby Walsh (b. 1964), a professor of artificial intelligence at the University of New South Wales, once gave his father a program that perfectly plays the game of Connect Four®. His dad, who had previously loved playing the game, remarked that the program took the fun out of the game, and Walsh had to agree. When smartphones become superior to humans in virtually all games and creative endeavors, from music composition to novel writing, what effect will that have on the collective psyche of humanity?

Connect Four is played by two people using discs (yellow vs. red) on a vertical board with seven columns and six rows. As the discs slide down a column to the bottommost open grid position, the objective is to be the first player to form a line of four adjacent discs of one's own color (horizontally, vertically, or diagonally). The game is reminiscent of a tic-tac-toe game, but with gravity influencing the pieces. Of course, Connect Four is much more complicated than tic-tac-toe: if one considers all possible game boards filled with 0 to 42 discs, there are a walloping 4,531,985,219,092 positions. In fact, the number of possible positions after n discs have been played on a standard 7 × 6 board grows as thus (where n = 0, 1, 2, 3, . . .): 1, 7, 56, 252, 1260, 4620, 18480, 59815, 206780, 605934, 1869840, 5038572, 14164920, 35459424, 91871208, 214864650, 516936420,

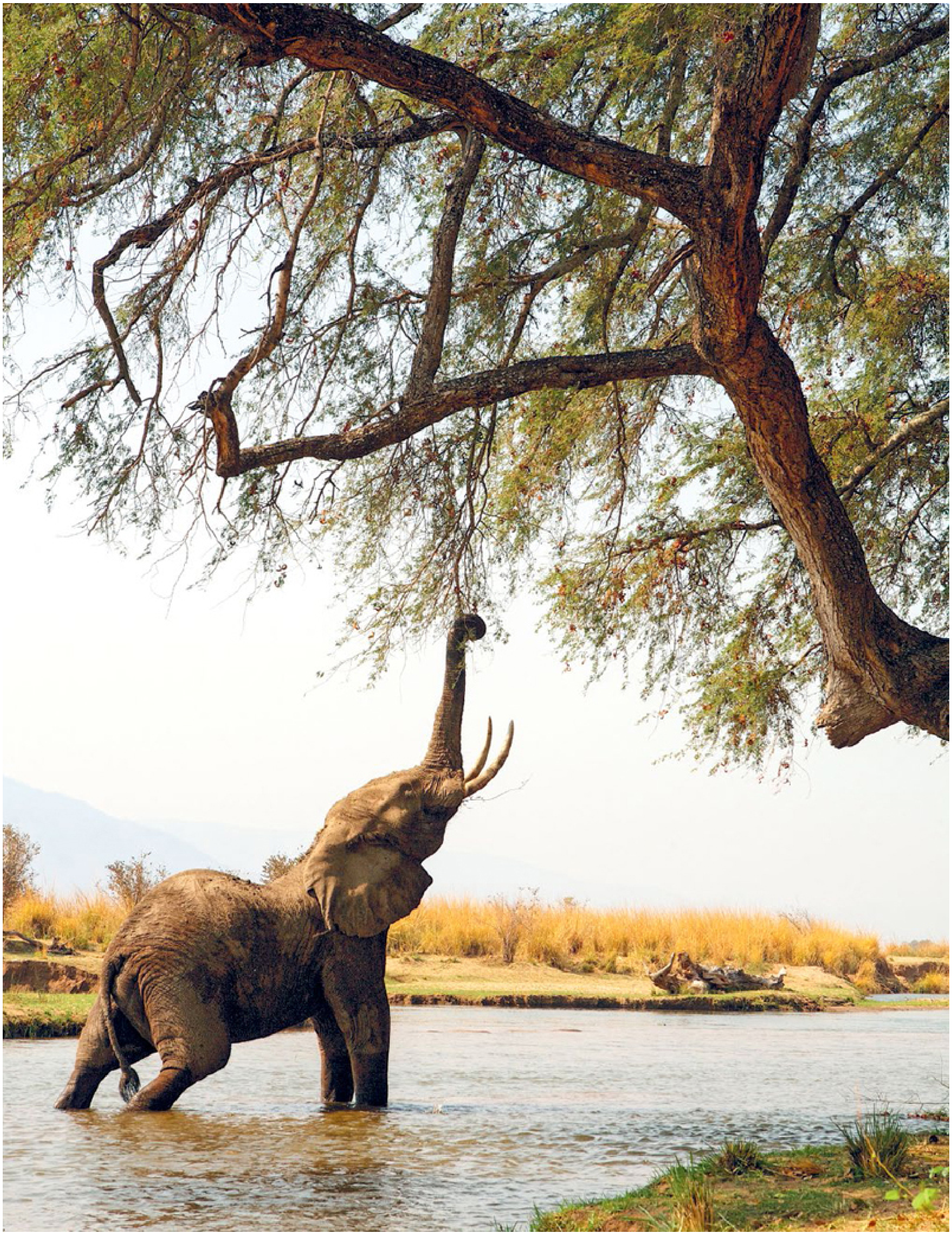
1134183050, 2546423880, 5252058812, 11031780760, 21406686756, 42121344720, 76871042612 . . .

On October 1, 1988, computer scientist James D. Allen finally "solved" Connect Four—that is, he devised an algorithm that could predict the outcome of moves (win, lose, or draw) from every possible position, assuming players play perfectly from that point on. Two weeks later, the game was solved independently by computer scientist Victor Allis, who employed an AI approach with nine strategies. As a result, we now know that Connect Four can always be won by the first player with perfect play.

There is much room for further research on variations of Connect Four. For example, imagine playing on boards that are wrapped into a cylinder, or playing on boards with different grid sizes, additional colors, and more than two dimensions. The number of possible positions and outcomes would be mind-boggling.

SEE ALSO Tic-Tac-Toe (c. 1300 BCE), Othello (1997), Solving the Game of Awari (2002), Quackle's Scrabble Win (2006), AlphaGo Go Champion (2016)

Intelligence in life forms clearly does not revolve around playing games like chess. In the paper “Elephants Don’t Play Chess,” Rodney Brooks argues for a different focus in exploring AI.



1990

"ELEPHANTS DON'T PLAY CHESS"

"There is an alternative route to Artificial Intelligence that diverges from the directions pursued under that banner for the last thirty some years," wrote roboticist Rodney Brooks (b. 1954) in his widely cited 1990 manifesto titled "Elephants Don't Play Chess." He continued: "The traditional approach has emphasized the abstract manipulation of symbols, whose grounding in physical reality has rarely been achieved. We explore a research methodology which emphasizes ongoing physical interaction with the environment as the primary source of constraint on the design of intelligent systems."

One of several points made by Brooks in his paper is that all around us are intelligences far removed from chess playing, in the form of intelligent animals (e.g., elephants) and even insect swarms. Brooks believed that AI research should move, at least in part, from a focus on classical AI involving rules, symbol manipulation, and search trees to one based more on considerations that include sensorimotor coupling with the environment (e.g., feedback between senses and motion-producing mechanisms), vision-movement coordination, and other forms of direct physical interaction with the real world.

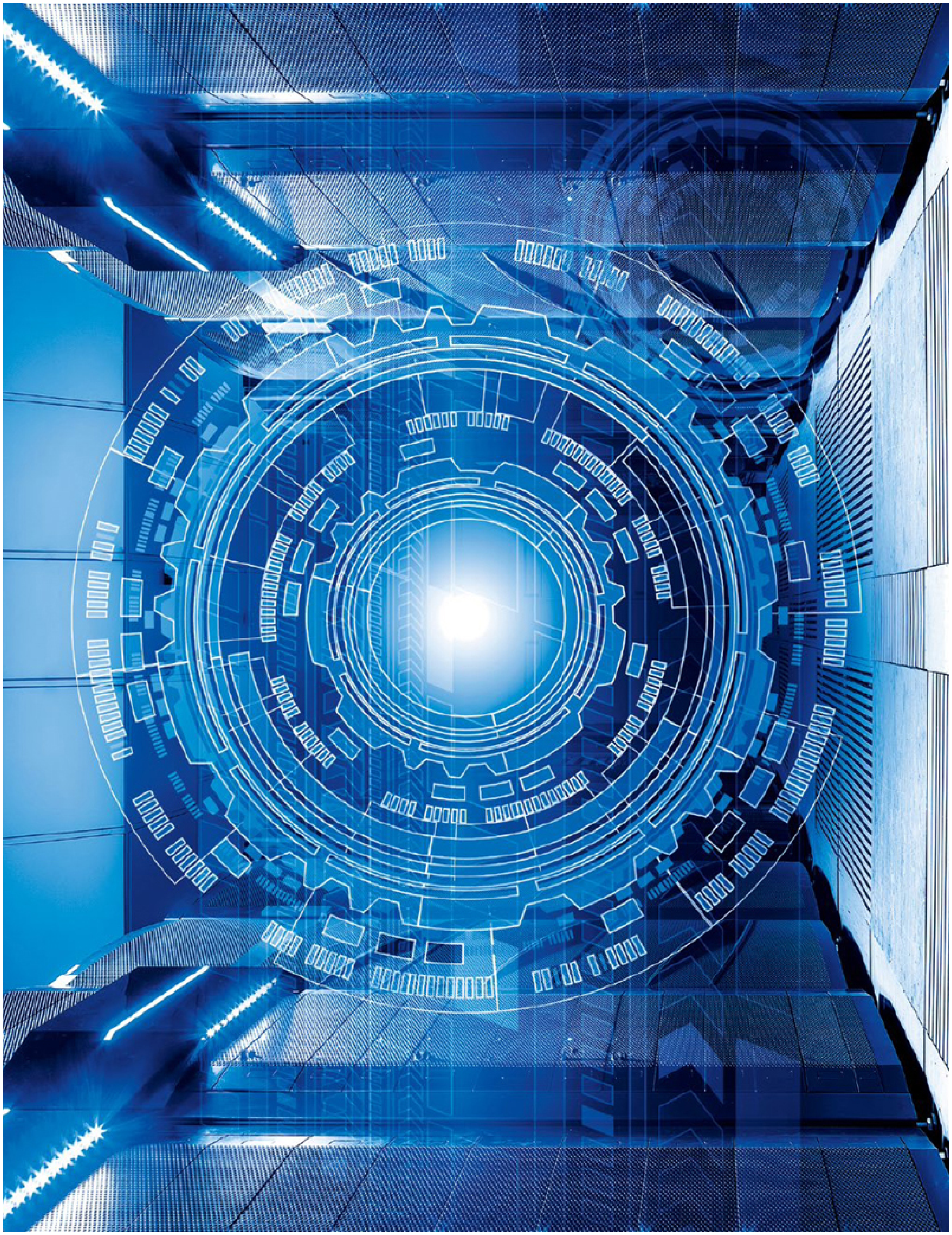
Brooks's paper concludes with a fascinating sampling of AI robots that are reminiscent of bundles of sensory systems that react to the environment. For Brooks, a physical body is considered to be crucial for the intelligence that interests him—in AI creatures that can solve

problems concerning locomotion, grasping, and navigating through the world. These behavior-based AI systems need not always "understand" the functioning of their separate behavioral units for their intelligence. For example, Brooks achieves intriguing behavior in his robots with simple rules involving avoiding static and moving obstacles, coupled with a "desire" to randomly wander and to look for distant places.

High-level behavior emerges from a set of simple interactions with the environment. On this environment topic, *Natural Computing* authors Dennis Shasha and Cathy Lazere note: "In the brief history of space travel, it has been easier to build a computer program to guide a spacecraft to Mars than to build a robot able to navigate over rough terrain with anything like the skill of a billy goat."

SEE ALSO " Darwin among the Machines" (1863), Turing Test (1950), Licklider's "Man-Computer Symbiosis" (1960), Swarm Intelligence (1986), Moravec's Paradox (1988)

The possible risks of highly advanced AI computer programs have led researchers to speculate about how to build AI enclosures that could confine or isolate such entities.



1993

LEAKPROOF "AI BOX"

As discussed in the entry "Intel-ligence Explosion," some scientists have expressed concerns that once AIs become sufficiently intelligent, such entities could repeatedly improve themselves so as to pose a possible threat to humanity. This runaway AI growth is sometimes referred to as the *technological singularity*. Of course, such entities might also be extremely valuable to humans; but possible risks have led researchers to speculate about how to build *AI boxes* that could confine or isolate such entities, if the need arises. For example, the hardware running the software for such entities might act as virtual prisons that are not connected to communications channels, including the Internet. The software could also be run on a software virtual machine within another virtual machine in order to increase the isolation. Of course, complete isolation would be of little value, since this would prohibit learning from—or observing—a superintelligence.

Nevertheless, if the AI superintelligence is sufficiently advanced, might it still be able to make contact with the outside world, or with various people who serve as gatekeepers, through unusual means, such as by altering the processor cooling-fan speeds to communicate via Morse code or by making itself so valuable that theft of the box is likely? Perhaps such an entity could be exceedingly convincing by offering its human gatekeepers bribes in order to coax them to allow more communication or replication to

other devices. This bribing may seem farfetched today, but who knows what wonders the AI could offer, including cures for diseases, fantastic inventions, melodies that enthrall, and multimedia visions of romance, adventure, or bliss.

Author Vernor Vinge (b. 1944) argued in 1993 that for superhuman intelligences, “confinement is intrinsically impractical. For the case of physical confinement: Imagine yourself locked in your home with only limited data access to the outside, to your masters. If those masters thought at a rate—say—one million times slower than you, there is little doubt that over a period of years (your time) you could come up with ‘helpful advice’ that would incidentally set you free.”

SEE ALSO “Darwin among the Machines” (1863), *Rossum’s Universal Robots* (1920), *Giant Brains, or Machines That Think* (1949), Intelligence Explosion (1965), Living in a Simulation (1967), Paperclip Maximizer Catastrophe (2003)

In the 1950s, IBM scientist Arthur Samuel became well known for his adaptive checkers program that learned by playing games against modified versions of itself. For what other games will AI players soon exhibit superhuman skills?

1994

CHECKERS AND AI

Checkers is played on an 8 × 8 board, with players who take turns trying to capture each other's pieces by hopping over them. In the 1950s, IBM scientist Arthur Samuel (1901–1990) became well known for his adaptive checkers program that learned by playing games against modified versions of itself. A famous milestone in the history of checkers-playing AI is Chinook, which in 1994 became the first computer program to win the world championship title against humans.

Chinook was developed by a team led by Canadian computer scientist Jonathan Schaeffer (b. 1957); it made use of both a library of opening moves played by grandmasters and an algorithm that, by 1992, searched checkers moves to an average minimum search depth of nineteen ply (one ply equals one move by one player). It also used an endgame database for all positions with eight pieces or fewer, and a useful move-evaluation function.

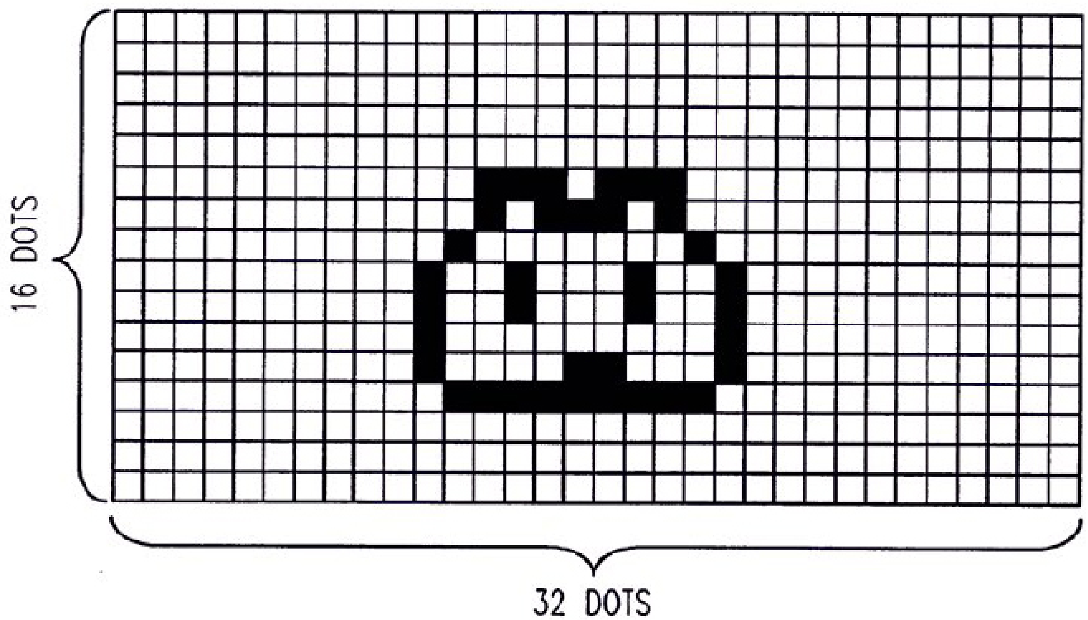
Before the famous 1994 man-machine checkers match, Marion Tinsley (1927–1995), widely considered the best checkers player who ever lived, declared "I have a better programmer than Chinook. His was Jonathan, mine was the Lord." Alas, after six games, which were all draws, Tinsley complained of abdominal pains and had to stop playing. He died a few months later of pancreatic cancer. Chinook was declared the winner by default.

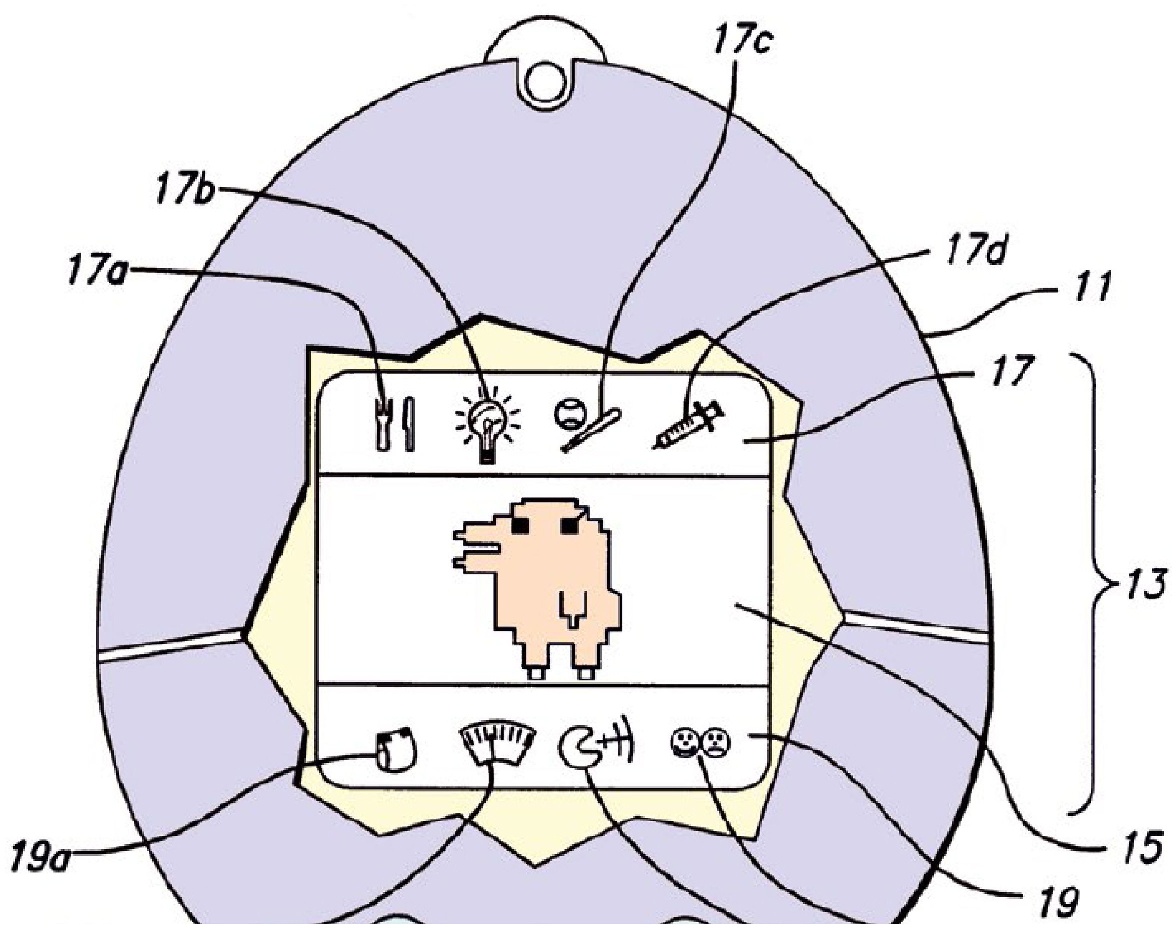
In 2007, Schaeffer and colleagues used computers to finally prove that checkers, when played perfectly, is a no-win game. This means that checkers resembles tic-tac-toe—a game that also can't be won if both players make no wrong moves. Schaeffer's proof was executed by hundreds of computers over eighteen years that eventually showed that it was theoretically possible to build a machine that will never lose to a human.

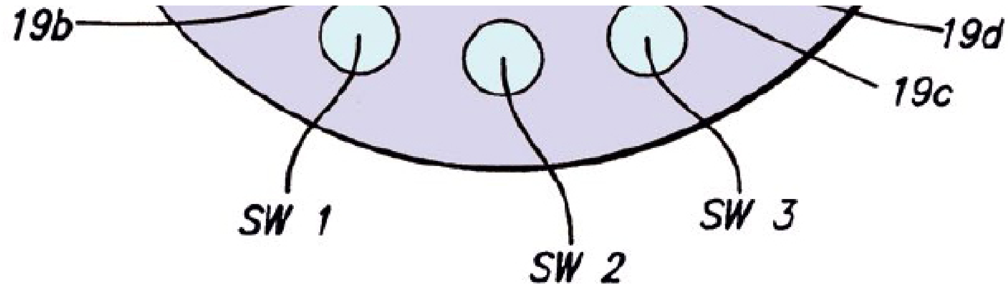
To "solve the game of checkers," the research team considered 39 trillion arrangements with ten or fewer pieces on the board and then determined if either of the two players wins. The team also used a specialized search algorithm to study the start of the game—specifically to see how initial moves "funneled" into the ten-checker configurations.

SEE ALSO Tic-Tac-Toe (c. 1300 BCE), Mechanical Turk (1770), Machine Learning (1959), Backgammon Champion Defeated (1979), Deep Blue Defeats Chess Champion (1997), Othello (1997)

Figure from US Patent 6,213,871, titled “Nurturing simulation apparatus for virtual creature,” issued to Akihiro Yokoi. The Tamagotchi form of artificial life, harbored in an egg-shaped device, started to become popular in 1997.






19b
19d
19c
SW 1
SW 2
SW 3

1996

TAMAGOTCHI

The Tamagotchi® artificial life-form, harbored in a tiny portable device, was among the first virtual pets to gain worldwide attention for children and adults. Amazingly, when it reached the United States at the FAO Schwartz toy store in 1997, 30,000 units were sold in three days. In a year, the AI was selling in more than eighty countries, producing revenues of more than $160 million. It also led to research into the "Tamagotchi Effect," which describes the emotional attachments people often develop with lifelike but inanimate forms. Parents also often wondered how to best handle the strong emotional upheavals of children when their virtual pets inevitably died. In the Japanese version, a dead pet would be represented by a ghost and headstone, and US versions might show an angel. Japanese toy-maker Bandai actually decided to open a virtual cemetery on the Internet for deceased pets.

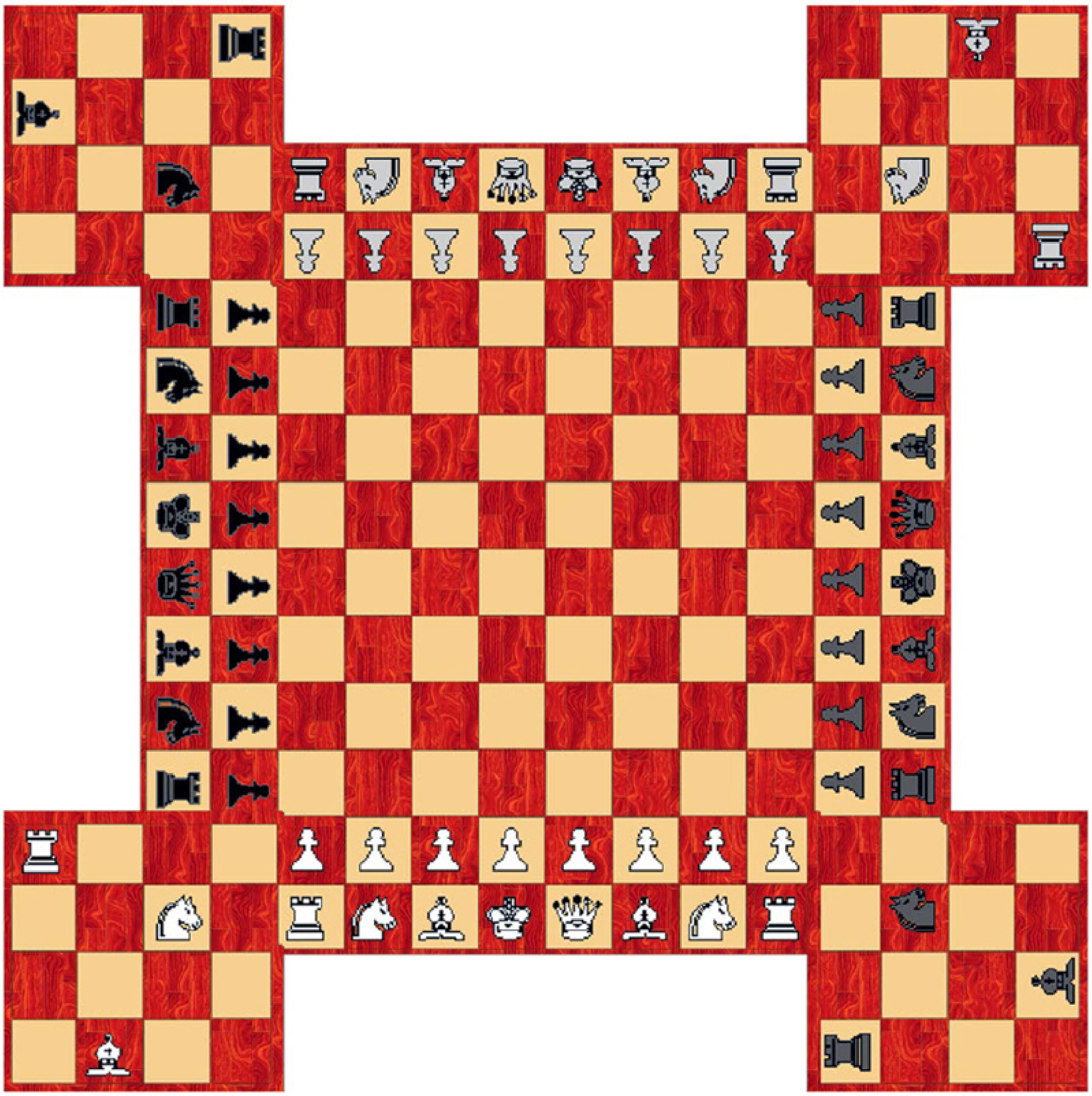
Developed by Bandai employee Aki Maita (b. 1967) and toy designer Akihiro Yokoi (b. 1955) in Japan, the Tamagotchi was first released in 1996. The software was housed in an egg-shaped object with an interface that consisted of just three buttons. A creature—displayed on a small, low-resolution screen—would start as an egg and then develop according to care the player provided. For example, owners would need to "feed" the pet, which could become sick if not cared for properly. Using infrared communications, two owners could link their toys to form friendships. Also, the device

would beep to request attention from the owner, and children often took their pets to school because the creatures could die in a few hours if they did not receive care. The ensuing distractions in the classroom caused some schools to ban Tamagotchis.

These simple forms of virtual life raised many questions, as children sometimes treated them as if they were actually alive. What is considered a healthy relationship between children and such entities, and how would this evolve in the coming years? Does the Tamagotchi have some form of intelligence, given that it “perceives” its environments through its infrared sensors and buttons, reacts, and requires socialization when it “feels” lonely? As advanced virtual pets are developed for the elderly for companionship, are possible risks outweighed by the rewards?

SEE ALSO Living in a Simulation (1967), Artificial Life (1986), AIBO Robot (1999)

What new versions of chess will AI entities or humans invent that may introduce new challenges for humans and machines? Shown here is Fortress Chess, played in Russia in the 18th and 19th centuries. Four players (AIs and/or humans) are represented here as black, white, dark grey, and light grey.



1997

DEEP BLUE DEFEATS CHESS CHAMPION

Vladimir Kramnik (b. 1975), the undisputed World Chess Champion from 2006 to 2007, once told reporters, "I am convinced, the way one plays chess always reflects the player's personality. If something defines his character, then it will also define his way of playing." Would this "personality" also be reflected in an AI's style of play?

For decades, technologists had considered chess a kind of measuring stick for artificial intelligence, being a game that required strategy, detailed reasoning, logic, foresight, and—at least for human players—guile. After many years of debating as to when a machine might defeat a reigning world chess champion, it finally happened in 1997: IBM's Deep Blue computer defeated Russian world chess champion Garry Kasparov (b. 1963) in a six-game match. After Game 5, Kasparov had become so discouraged that he explained: "I'm a human being. When I see something that is well beyond my understanding, I'm afraid."

Using special-purpose hardware, the 1997 version of Deep Blue was capable of evaluating 200 million chess positions per second, and it typically searched to a depth of around six to eight future moves, if not far more. Deep Blue's strategy could also take into account a large database of past grandmaster games, and it used endgame databases that included chess positions with five or fewer pieces.

The dream of chess machines stretches far back in time. The Mechanical Turk chess robot, created by Hungarian inventor Wolfgang von Kempelen in 1770, played a strong game of chess, but it also employed a human who hid inside the machine. In 1950, computer scientist Alan Turing and mathematician David Champernowne (1912–2000) designed a computer program, known as the “Turbochamp,” to play chess. However, because no computer was available to actually run the algorithm, Turing simulated a computer by manually consulting the algorithm during a testing phase.

In 2017, the program AlphaZero beat world-champion chess-playing computer programs, having taught itself how to play in less than a day! The program used machine learning, starting from random play, and was given no domain knowledge except the game rules.

SEE ALSO Mechanical Turk (1770), Checkers and AI (1994), AlphaGo Go Champion (2016)

Close-up of Othello playing pieces, showing that the discs may be flipped to turn from black to white, or vice versa, according to the rules of the game.

1997

OTHELLO

Because of the precise rules of games and the ability to determine a winner in the playing process, games have been a popular testing arena for AI researchers. Also, the AI system can often play millions of games against itself or other AI players so as to improve and gain insight. One interesting game to which AI has been successfully applied is Othello, also called Reversi, which was mentioned in an 1886 edition of *The Saturday Review* and dates much farther back in time.

The game of Othello is played on an 8 × 8 grid with disks that are white on one side and black on the other. During a turn, a player places a disk with the player's color facing up.

Imagine you are playing black, and you place your disk at one end of a line of white pieces with another black disk at the other end. In that scenario, the intermediate white disks are turned over to black. In other words, discs are flipped by a play that sandwiches them between opposite color discs, and the player with the majority color at the end of the game is the winner.

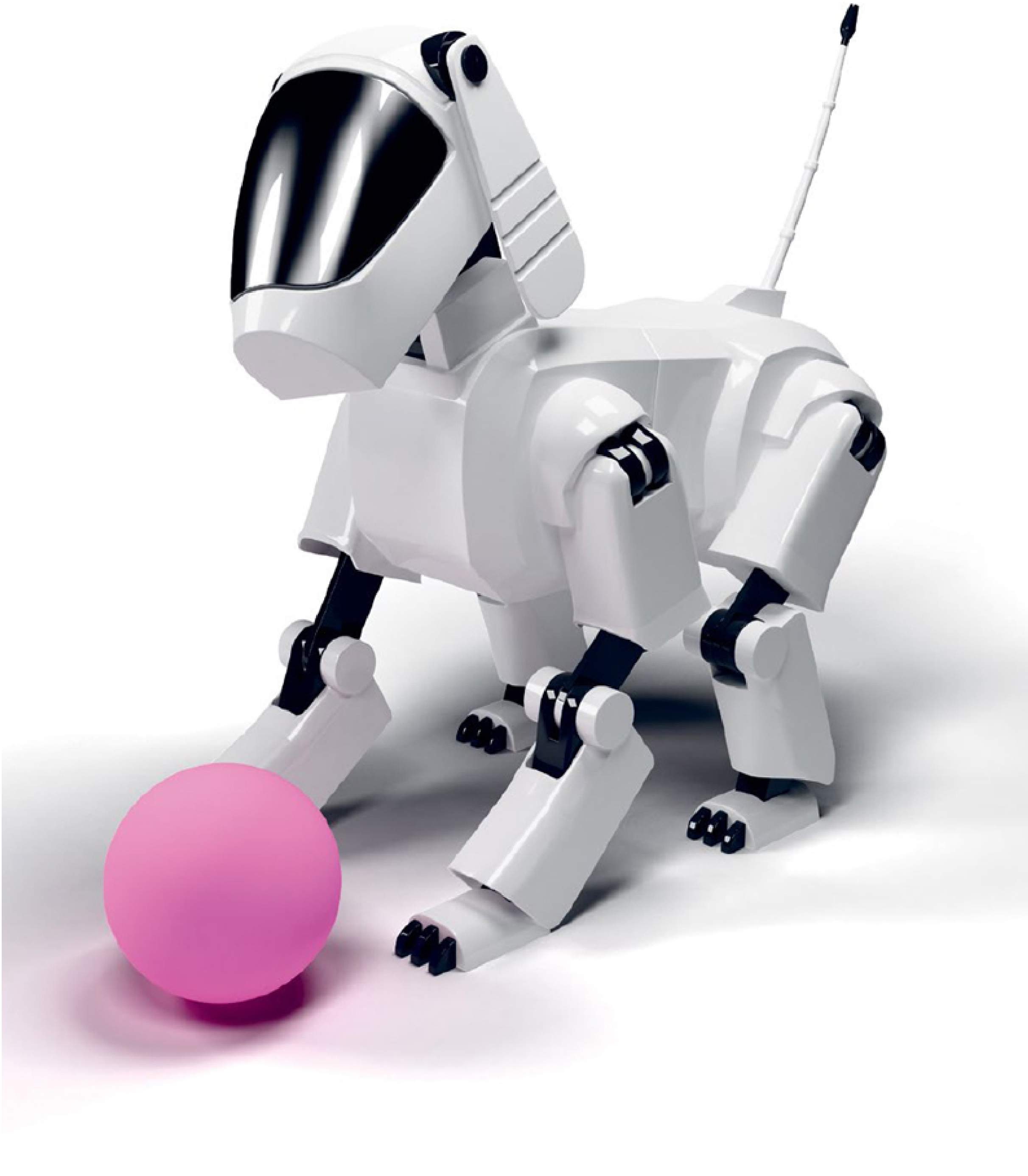
One reason the game is so challenging for humans to visualize is that, unlike chess and checkers, the colors of Othello pieces may keep changing. Since around 1980, Othello computer programs have easily defeated professional players. In 1997, Logistello, a computer program created by computer scientist Michael Buro, defeated the human champion Takeshi Murakami by the score of 6

games to 0. At the time, Murakami, a 32-year-old English teacher from Tokyo, seemed surprised at his defeat, saying that at least one Logistello move was “unfathomable” because it was so different from a move a human would make.

The number of legal positions in Othello is at most 10^{28}. Even today, Othello is considered an “unsolved game” in the sense that no one has proved what the outcome of a game would be if both sides played perfect games.

SEE ALSO Tic-Tac-Toe (c. 1300 BCE), Mastermind (1970), Checkers and AI (1994), Deep Blue Defeats Chess Champion (1997), AlphaGo Go Champion (2016)

As robot dogs and other pets become more advanced in the coming years, will there come a time when they are nearly indistinguishable from actual animals or outnumber actual living pets in our homes?



1999

AIBO ROBOT

The dog-like robot AIBO, introduced by the Sony Corporation in 1999, was one of the world's first mass-market sophisticated consumer robots for entertainment. Enjoyed by children and adults, AIBO also was used in AI education and for research purposes because it included a vision system and articulators in a relatively inexpensive package. The robots have also been used in RoboCup® autonomous soccer competitions, many of which can be viewed on YouTube, showing how the dogs search for the ball and move it toward the goal.

The AIBO, which takes its name from the Japanese word for "pal," could respond to numerous commands and made use of various sensors, including touch sensors, a camera, a range-finder, and a microphone. More sensors and actuators (for the moveable legs, neck, and other parts) were added in successive models of the robot. Some versions could autonomously recharge their batteries at a charging station. Software gave the robots a personality and the ability to walk and respond to the environment. Each robot learned slightly different behaviors as people interacted with it.

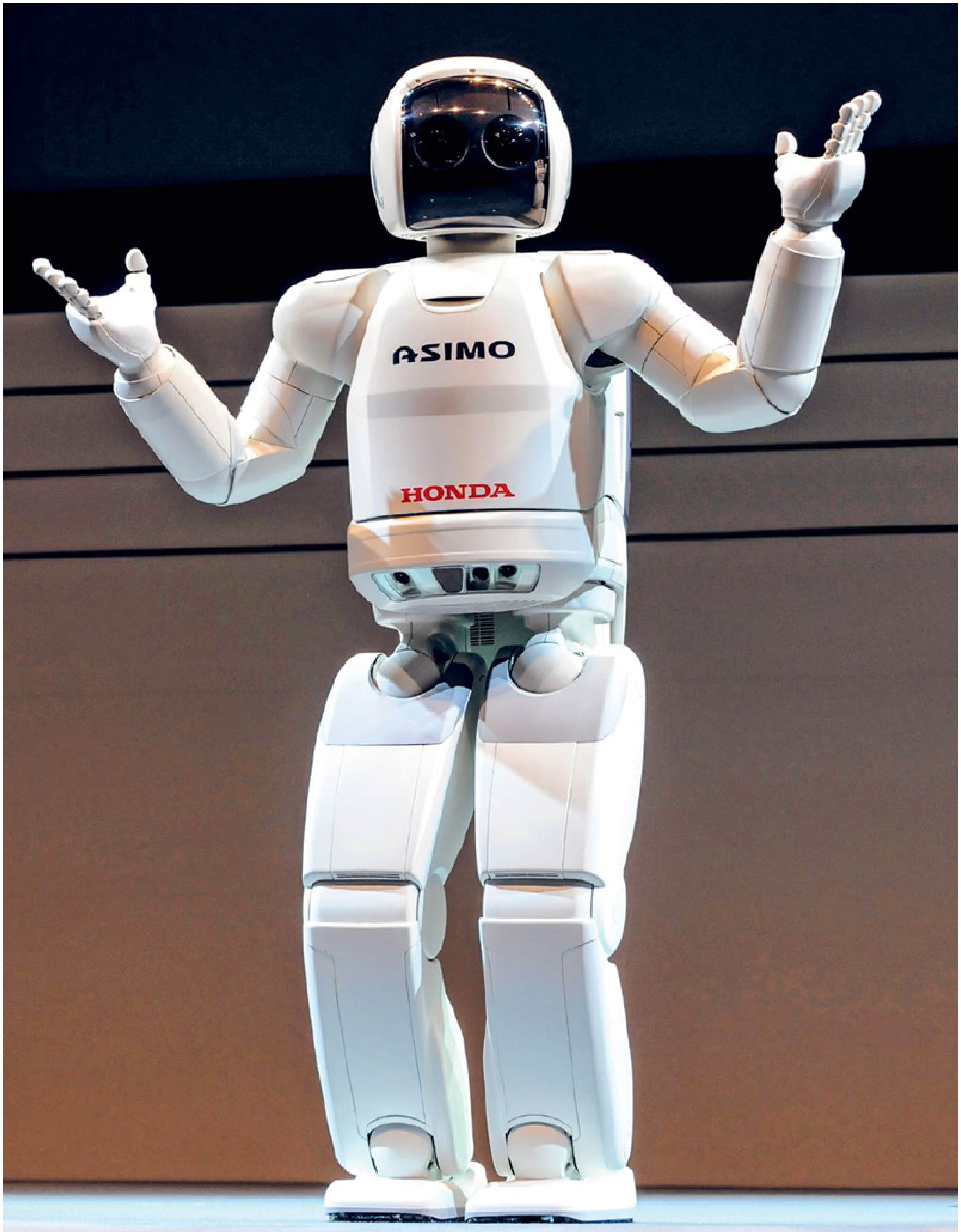
Fascinating studies have been conducted using the AIBO and other artificial pets for patients with dementia, and the studies suggest that robotic pets could be beneficial for companionship and stimulation. More advanced pets may one day be even more helpful to people during any period of cognitive decline. Other studies have

considered AIBO–human relationships and found that a significant number of owners reported that the AIBO has feelings, even though they understood that the dog was mechanical and not alive. Psychologists continue to ponder the implications of autonomous systems that fool people into thinking they have real emotions and far more capabilities than they actually have.

In 2017, Sony announced a new generation of AIBO with many more actuators to provide smoother, more natural movements. These fourth-generation models have improved face recognition abilities, more features involving mobile Internet access, and more sophisticated abilities to adapt, learn, and respond to the environment.

SEE ALSO de Vaucanson's Duck Automaton (1738), *Electric Bob's Big Black Ostrich* (1893), Face Recognition (1964), Shakey the Robot (1966), Tamagotchi (1996), ASIMO and Friends (2000)

A humanoid robot. If you had a robot butler residing in your home, would you prefer the butler to look human, or would a more robotic appearance be more comfortable for you?


ASIMO
HONDA

2000

ASIMO AND FRIENDS

The history of real-world robots has had several notable milestones, just a few favorites of which are listed here. Three-wheeled “tortoises,” developed in 1949 by British neurophysiologist William Walter (1910–1977), could autonomously explore the environment using various sensors. In 1961, Unimate, created by American inventor George Devol (1912–2011), became the world’s first industrial robot and was employed on the General Motors automobile assembly line. In 1973, Japan’s WABOT-1—the world’s first full-scale, humanoid, intelligent robot—required 45 seconds for each step. In 1989, MIT showed off the six-legged robot insect named Genghis, developed by roboticist Rodney Brooks (b. 1954), which employed simple logic rules in order to walk and explore. In 1998, Tiger Electronics released the owl-like Furby®, which sold over 40 million units in a few years. Although a very simple robot, Furby produced “Furbish” speech output that transformed into English over time, giving the impression that it could learn language just as humans do. Finally, BigDog, a four-legged robot created in 2005 by Boston Dynamics and partners, was notable for its ability to walk across a variety of difficult terrains.

Perhaps one of the most iconic of modern real-world robots has been Advanced Step in Innovative Mobility (ASIMO®), introduced by the Honda Motor Company in 2000 to significant mainstream fanfare. The humanoid robot was 4 feet, 3 inches tall (130 cm) and,

using its internal cameras and various sensors, was capable of autonomous navigation by walking. ASIMO could recognize gestures, faces, and sounds, and it could also grasp objects.

Our complex future will surely have a continuing need for AI, and robots will play an increasing collaborative role with humans. Perhaps someday robots like ASIMO will assist the elderly or infirm. However, as cyberneticist Norbert Wiener cautions, “The world of the future will be an even more demanding struggle against the limitations of our intelligence, not a comfortable hammock in which we can lie down to be waited upon by our robot slaves.”

SEE ALSO Elektro the Moto-Man (1939), Shakey the Robot (1966), AIBO Robot (1999), Roomba (2002), AI on Mars (2015)

The Adventures of Pinocchio was one of the inspirations for Stanley Kubrick during the development of *A.I*. Pinocchio, the wooden puppet, dreamed of becoming a human boy.

2001

SPIELBERG'S *A.I. ARTIFICIAL INTELLIGENCE*

The film *A.I. Artificial Intelligence*, directed by Steven Spielberg (b. 1946), is a provocative drama that makes us question the future of AI and the capabilities of an AI android boy named David, given as a gift to a mother because she misses her human son. The movie is based on the 1969 story "Supertoys Last All Summer Long" by English author Brian Aldiss (1925–2017). Although the film was not released until 2001, its development began in the 1970s, when director Stanley Kubrick acquired the film rights.

Much of the film follows David after his separation from his mother and his attempt to return to her. He is accompanied on his quest by Teddy, an AI teddy bear, as he seeks a Blue Fairy (inspired by the Disney film *Pinocchio*), who David believes can turn him into a "real" human being. In one part of the journey, David's engineer creator explains to David that the "Blue Fairy is part of the great human flaw to wish for things that don't exist, or the greatest single human gift—the ability to chase down our dreams. And that is something no machine has ever done until you."

Viewers of the film debated whether an android could truly love, with film critic Roger Ebert (1942–2013) writing that an android is merely "a puppet with a computer program pulling its strings." Toward the end of the film, after two thousand years in suspended animation, David encounters spindly alien AIs that have evolved

from androids like him and who have taken a particular interest in David, because he is the last being who has actually ever encountered humans. They allow him to spend one last day with a clone of his long-dead human mother in a virtual dream-state. Reflecting on the sad but thought-provoking film, *Empire of Dreams* author Andrew Gordon writes: “We measure ourselves against the robot, trying to define what makes us human, fearing that even as we become more robotic, our creations become more human and may ultimately surpass or replace us. . . . A robot has become ‘human’ by attaining the ability to dream, yet dreams of the human are all that is left of humanity.”

SEE ALSO “The Artist of the Beautiful” (1844), Transhumanism (1957), *Blade Runner* (1982)

Awari has fascinated researchers in the field of artificial intelligence. In 2002, computer scientists calculated the outcome for all 889,063,398,406 positions that can occur in the game, and proved that Awari must end in a draw for perfect players.

2002

SOLVING THE GAME OF AWARI

Researchers studying artificial intelligence have invested great effort developing game-playing programs—both as a test of AI strategies and to push the limits of software and hardware. One interesting example in the history of game playing involves Awari, a 3,500-year-old African board game. Classified as a count-and-capture game, Awari is a member of a set of strategy games called *mancala* and has various different names in different countries.

The Awari board consists of two rows of six cup-like hollows—with four markers (beans, seeds, or pebbles) in each hollow—where one row belongs to each player. On a player's turn, the player chooses one of his or her six cups, withdraws all seeds from that cup, and drops one seed in each cup counterclockwise from this cup. The second player then takes the seeds from one of the six cups on his or her side and does the same. When a player drops his last seed into a cup on the opponent's side containing only one or two seeds (making a total of two or three seeds), that player captures all the seeds in this cup. The same player also takes any seeds in cups immediately before the emptied cup if they now also total two or three. Players can take seeds only from their opponent's side of the board, and the game ends when one player has no seeds left in the cups on his or her side. Whoever captures the majority of seeds wins.

Awari has been of immense attraction to researchers in the field of artificial intelligence; but until 2002, no one knew if the game was like tic-tac-toe in which perfect plays from the beginning always ended a game in a draw. Finally, computer scientists John W. Romein (b. 1970) and Henri E. Bal (b. 1958) of the Free University in Amsterdam wrote a computer program that calculated the outcome for all 889,063,398,406 positions that can occur in the game, proving that Awari must end in a draw with perfect play by both players. The massive computation required about 51 hours on a computer cluster with 144 processors.

"Did we ruin a perfectly fine game?" ask Romein and Bal. "We do not think so. Connect-4 was solved as well, and people still play the game. The same holds for other solved games."

SEE ALSO TicTac-Toe (c. 1300 BCE), Mastermind (1970), Backgammon Champion Defeated (1979), Connect Four (1988), Checkers and AI (1994), Othello (1997)

Modified versions of robotic vacuum cleaners have been hacked by hobbyists to alter their functionality, causing the robots to map rooms, draw Spirograph®-like patterns, battle one another, perform surveillance with cameras, and more.

2002

ROOMBA

A year after the Roomba® robotic vacuum cleaner was introduced to the world by iRobot Corporation, journalist Monte Reel wrote: “How did it happen that the robotic revolution is kicking off with the Roomba, an automated floor vacuum cleaner that will mesmerize—or maybe terrify—your pet cat? It was a fortuitous meeting of brain power, ambition, and a distaste for housework on the part of iRobot engineers, whose backgrounds range from artificial-intelligence research to the design of unmanned extraterrestrial vehicles.”

The Roomba autonomic vacuum cleaner, which appeared on the market in 2002, contains various sensors that enable it to detect dirt on the floor, sense stairways that might cause it to fall, and change its travel direction when encountering objects. The 900 series also featured a camera that helped the navigation software ensure that the device efficiently covered the floor area. When the battery in the Roomba is low, the vacuum seeks a charger by using its infrared beacons. Acoustic dirt sensors help it detect certain kinds of dirty spots that might need special attention. In the past, the Roomba has relied on spiral cleaning paths, random walks, and other methods for covering a floor.

A variant of the LISP programming language was used to write much of the software for Roomba. Today, various hackers enjoy modifying the device, using its Roomba Open Interface (ROI) for

other purposes, like turning the Roomba into a painter or using it for surveillance.

The Roomba is a representative of a wider class of domestic robots used for household chores. Of course, as these devices become more sophisticated, privacy concerns will arise. Already, debates have raged regarding the maps of a user's home that iRobot could conceivably share with any businesses interested in estimating aspects of a user's home or lifestyle, and perhaps the police could even interrogate such robots during crime investigations.

SEE ALSO Shakey the Robot (1966), Autonomous Vehicles (1984), ASIMO and Friends (2000), AI on Mars (2015)

Even if programmed with a beneficent task, such as efficiently manufacturing paperclips, what if the AI then decided to convert as much of the planet as possible into a paperclip manufacturing facility?

2003

PAPERCLIP MAXIMIZER CATASTROPHE

Even if highly intelligent and capable AIs are given useful goals, it is nevertheless possible to have dangerous results in the future. One famous example, discussed in 2003 by philosopher and futurist Nick Bostrom, is the horror of the Paperclip Maximizer. Imagine a future in which an AI system supervises a set of factories producing paperclips. The AI is given a mission of producing as many paperclips as possible. If the AI is not sufficiently constrained, one might imagine that it could optimize its goal, first by operating the factories at maximum efficiency, and then devoting more and more resources to the task until vast regions of land and more factories were devoted to making paperclips. Eventually, all the available resources on Earth could be converted to this task, followed by all the relevant matter in the Solar System, turning everything into paperclips.

Although this scenario may sound implausible, it is intended to focus attention on the serious point that AIs may not have humanlike motives that we can truly understand. Even innocuous goals could prove dangerous if AIs gain the ability to evolve and create their successive improved machines, as discussed in the entry on Intelligence Explosion. How might humans ensure that AI goals, and their constituent mathematical reward and utility functions, remain stable and comprehensible through the decades and centuries? How

might useful “off switches” remain accessible? What if such reward circuits and software cause the AI to lose interest in the external world and devote its energy to maximizing a reward signal, much like taking a drug and dropping out of society? How will we deal with different countries and geopolitical regimes using different reward functions for their own AIs?

Another famous example, attributed to AI expert Marvin Minsky, is the Riemann hypothesis catastrophe, in which we imagine a superintelligent AI system with the goal of solving this difficult and important hypothesis in mathematics. Such a system might devote more and more computing resources and energy to the task, taking over and creating ever-improved systems at the expense of humanity.

SEE ALSO “Darwin among the Machines” (1863), Intelligence Explosion (1965), Leakproof “AI Box” (1993)

In Scrabble, each tile has a single letter, which is annotated with its value in the range 1 to 10, based on its frequency of use in the English language. Vowels are worth only one point.

2006

QUACKLE'S SCRABBLE WIN

"Chess had its Deep Blue," writes com-puting technology journalist Mark Anderson. *Jeopardy!* had its Watson. Baseball has its sabermetrics, as chronicled in the hit book and film *Moneyball*. In each game, data mining has upended the field of play." An exciting development involving AI applied to word games occurred in 2006, when the computer program Quackle defeated former world champion David Boys at Scrabble® in a tournament held in Toronto, Canada. When Boys lost three games out of five, he insisted "It's still better to be a human than to be a computer."

The Scrabble game was invented in 1938 by American architect Alfred Butts (1899–1993). When playing Scrabble, opponents place tiles on a game board consisting of a 15 × 15 array of squares. When playing the English version, each tile has a single letter, which is annotated with its value in the range 1 to 10, based on its frequency of use in the English language. For example, vowels are worth one point, whereas *Q* and *Z* are worth 10 points. As players alternate turns, they must add letters to the board in such a way that each row or column of letters always forms a word.

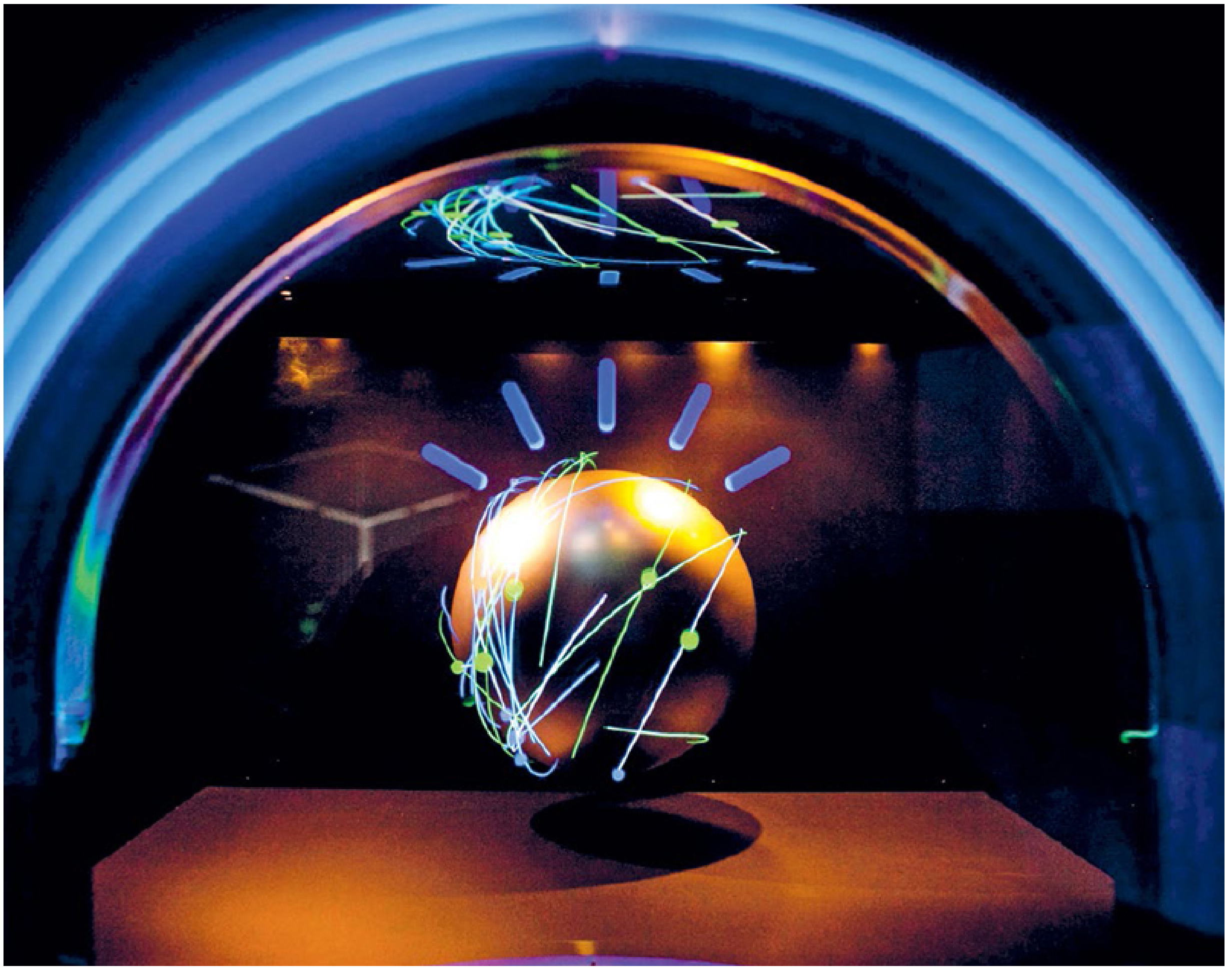
The game is actually quite complex, invol-ving more than just knowledge of words in a language. For example, strategy may involve anticipating what letters are still to be drawn as well as the utility of placing tiles on special squares on the board that may increase the tile's score. Scrabble is considered a game of imperfect

information, like poker, since an opponent's available tiles are hidden from view.

Quackle determines what tiles to play based on simulations that assess the board using an evaluation function. It was created by a team that included Jason Katz-Brown, who himself is one of the top-ranked Scrabble players in the world. As part of their fascinating research, developers had Quackle play itself numerous times to better understand the value of playing various words compared to other words a player might legally use during a turn, or with considerations of words that a player might use on the next turn.

SEE ALSO Connect Four (1988), Deep Blue Defeats Chess Champion (1997), Othello (1997), AlphaGo Go Champion (2016), AI Poker (2017)

The globe-like avatar representing IBM Watson is based on the IBM Smarter Planet icon. In the display employed for the *Jeopardy!* contest, colors and motions changed depending on game states and answer confidence levels.



2011

WATSON ON *JEOPARDY!*

Ken Jennings (b. 1974), the world champion of the game show *Jeopardy!*®, wrote about his contest with an AI entity called Watson: "When I was selected as one of the two human players to be pitted against IBM's 'Watson' supercomputer in a special man-vs.-machine *Jeopardy!* exhibition match, I felt honored, even heroic. I envisioned myself as the Great Carbon-Based Hope against a new generation of thinking machines. . . ."

Watson was a question-answering computer system that used natural-language processing, machine learning, information retrieval, and more to defeat the world champion in 2011 in a game involving general knowledge clues. What made the task particularly difficult—more difficult than chess playing—is that the computing system needed to provide an answer in just a few seconds, while considering the challenges and ambiguities of the English language, with clues that included puns, humor, riddles, cultural references, special contexts, and rhymes—something that humans consider instinctively.

To accomplish this task, Watson employed thousands of parallel processing units called cores, along with information such as the entire Wikipedia corpus, stored in its RAM memory (because access to spinning hard drives would be too slow during the competition). All information had to be stored locally, since Watson was not allowed access to the Internet during a match; in order to arrive at an answer,

the AI considered the results of numerous separate analysis algorithms at once. The more algorithms that found the same answer, the more likely it would be a correct answer. Watson continually scored different answers with a confidence level, and if the confidence was sufficiently high, Watson would provide the answer.

After his loss, Jennings wrote “But there’s no shame in losing to silicon. . . . After all, I don’t have 2,880 processor cores and 15 terabytes of reference works at my disposal—nor can I buzz in with perfect timing whenever I know an answer. My puny human brain, just a few bucks worth of water, salts, and proteins, hung in there just fine against a jillion-dollar supercomputer.”

SEE ALSO Natural Language Processing (1954), Machine Learning (1959), Deep Blue Defeats Chess Champion (1997), Quackle’s Scrabble Win (2006)

Example of a DeepDream artwork. The approach makes use of an artificial neural network that can seek and enhance patterns in images, with startling results.

2015

COMPUTER ART AND DEEPDREAM

According to essayist Jonathan Swift (1667–1745), "Vision is the art of seeing things invisible." This idea of seeing new patterns at the edges of art, science, and mathematics certainly applies to many categories of art produced with the aid of computers, algorithms, neural networks, and other forms of AI. Early explorations of computer art include the work of Desmond Paul Henry (1921–2004) and his bombsight analog computer-based drawing machines, starting around 1961. American engineer A. Michael Noll (b. 1939) became famous in 1962 for exploring random and algorithmic processes to produce visual art, and British-born artist Harold Cohen (1928–2016) created AARON—an AI computer-drawing program that could autonomously produce art—in 1968.

A more recent example of computer art includes the collaboration of many users with DeepDream, a 2015 computer-vision program created by Google engineer Alexander Mordvintsev and colleagues. The approach makes use of an artificial neural network (ANN) that can seek and enhance patterns in images—with startling results. To better understand DeepDream, consider that neural nets may be trained to classify and recognize features in an input image (e.g., chipmunks or stop signs) based on numerous "training" images. By running the neural network "in reverse," DeepDream looks for patterns in images and amplifies them in a manner somewhat reminiscent of when we gaze up at the clouds and begin to see

animal-like shapes. For artificial neural nets, each layer progressively extracts higher-level features—for example, the first layer might be sensitive to corners and edges, while layers closer to the output neurons may be examining complex features. Not only are the resulting pictures fascinating to study and brimming with detail, but they can provide a sense of the level of abstraction on which a particular ANN layer is working.

The resemblance of DeepDream artwork to hallucinations experienced by users of certain mind-altering drugs suggests that DeepDream might help researchers better understand how artificial neural networks relate to actual neural networks in the brain's visual cortex. Furthermore, the program could help illuminate how the brain attempts to find pattern and meaning.

SEE ALSO Computational Creativity (1821), Artificial Neural Networks (1943), Deep Learning (1965), *Cybernetic Serendipity* (1968)

In the future, Kelly believes that we will encounter "synthetic aliens," beings that we create, forcing us to deal with the same benefits and challenges that we would expect if we ever established contact with intelligent extraterrestrial aliens from other planets.

2015

"CALL THEM ARTIFICIAL ALIENS"

"The most important thing about making machines that can think is that they will think differently," writes Kevin Kelly (b. 1942), founding Executive Editor of *Wired* magazine. In his acclaimed 2015 essay, "Call Them Artificial Aliens," he noted that ". . . to solve the current grand mysteries of quantum gravity, dark energy, and dark matter, we'll probably need intelligences other than human. The extremely complex questions that will come after them may require even more distant and complex intelligences. Indeed, we may need to invent intermediate intelligences that can help us design yet more rarified intelligences that we couldn't design alone."

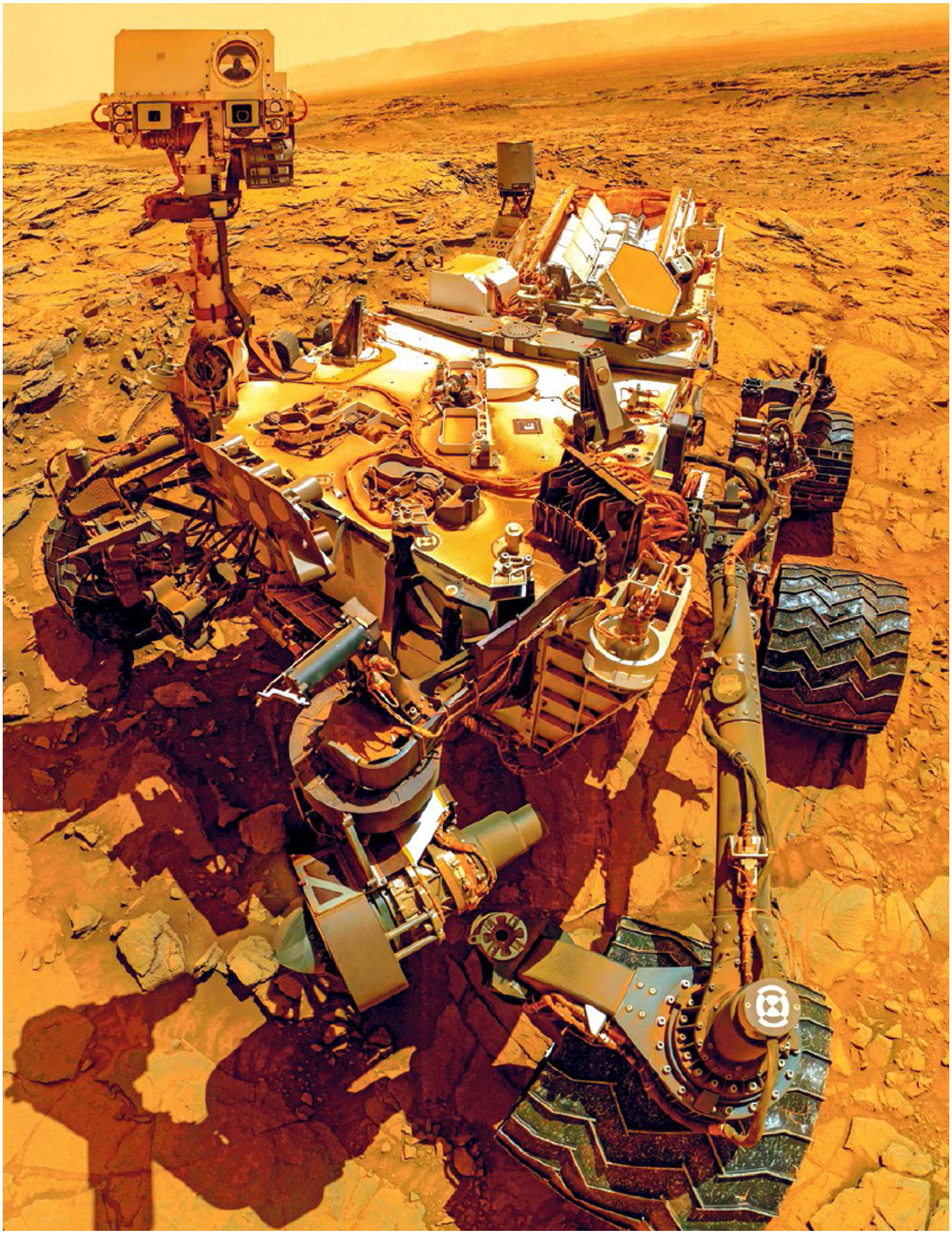
So deep and difficult will be questions in the future that they will require numerous different "species of minds" to solve them, along with new human skills to interface with such minds. Kelly concludes his essay by comparing thinking machines with aliens: "AI could just as well stand for Alien Intelligence. We cannot be certain that we'll contact extra-terrestrial beings . . . in the next 200 years, but we can be almost 100 percent certain that we'll have manufactured an alien intelligence by then. When we face those synthetic aliens, we'll encounter the same benefits and challenges we expect from contact with ET. They'll force us to reevaluate our roles, our beliefs, our goals, our identity."

We can hardly imagine an antelope understanding the significance of prime numbers, yet alterations of our own brains, along with the

development of useful AI interfaces, could admit a variety of profound concepts to which we are now totally closed. If the yucca moth, with only a few ganglia for its brain, can recognize the geometry of the yucca flower from birth, how much of our capacity is hardwired into our convolutions of cortex? Of course, there are likely facets of the universe we can never understand, just as an antelope could never understand calculus, black holes, symbolic logic, or poetry. There are thoughts we can never think, visions we can only glimpse. It is at this filmy, veiled interface between human reality and a reality beyond that we may find the numinous, which some may liken to dancing with the artificial gods.

SEE ALSO Searches for the Soul (1907), *Giant Brains, or Machines That Think* (1949), Intelligence Explosion (1965), Artificial Life (1986)

Self-portrait of *Curiosity* located at the foothill of Mount Sharp on Mars, October 6, 2015.



2015

AI ON MARS

Artificial intelligence and autonomy will play increasing roles in space exploration as robotic spacecraft and rovers need to make rapid, smart decisions, particularly when they are not in constant communication with Earthlings back home. As we reach further into the solar system, and even send rovers into the seas of faraway moons like Jupiter's Europa, communication delays can be particularly troublesome.

One recent example of limited but fascinating AI in outer space involves the NASA *Curiosity* rover, patrolling Mars to help determine whether the planet was ever capable of sustaining life, and to help us better understand Martian geology, climate, and radiation patterns. In 2015, *Curiosity* received an AI software upload called Autonomous Exploration for Gathering Increased Science (AEGIS) to help it with its tasks.

"Right now Mars is entirely inhabited by robots," says planetary scientist Raymond Francis, "and one of them is artificially intelligent enough to make its own decisions about what to zap with its laser." If *Curiosity* finds a particular surface feature of interest, it can vaporize a small portion with a laser and examine the resulting spectra to estimate the rock's composition. If warranted, it can also use its long arm, microscope, and X-ray spectrometer for a closer examination. In particular, AEGIS allows *Curiosity* to autonomously select a target rock and accurately pinpoint it with the laser—using computer vision

to examine digital images, looking for edges, shapes, size, brightness, and so forth. Journalist Marina Koren notes: “The software—just about 20,000 lines of code out of the 3.8 million that make *Curiosity* tick—has turned a car-sized, six-wheeled, nuclear-powered robot into a field scientist.”

Curiosity’s investigations may one day pave the way for human exploration, and systems like AEGIS will aid in this process by using machine learning and other AI approaches to detect anomalies. Given that there are periods of time when *Curiosity* is on the far side of Mars and cannot communicate with Earth and receive instructions, and given that communication with Earth can drain power from the rover, AI can be particularly useful when communication is limited or impossible.

SEE ALSO Shakey the Robot (1966), Autonomous Vehicles (1984), Roomba (2002)

The AlphaGo computer program defeated South Korean Lee Sedol, the first time a program beat a 9-dan professional player, without handicaps.

2016

ALPHAGO GO CHAMPION

"While the Baroque rules of Chess could only have been created by humans, the rules of Go are so elegant, organic, and rigorously logical that if intelligent life forms exist elsewhere in the universe, they almost certainly play Go," explained German-American chess and Go expert Edward Lasker (1885–1981).

Go is a two-player board game that originated in China possibly around 2000 BCE and spread to Japan, where it became popular in the thirteenth century CE. In the game, players alternately place black-and-white stones on intersections of a 19 × 19 playing board. A stone or a group of stones is captured and replaced with stones of the opposite color if it is tightly surrounded by stones of the opposing player, and the objective is to control a larger territory than one's opponent. Go is complex for many reasons, including its large game board, complex strategies, and huge numbers of variations possible in games. In fact, the number of potential legal board positions is far more than the number of atoms in the visible universe!

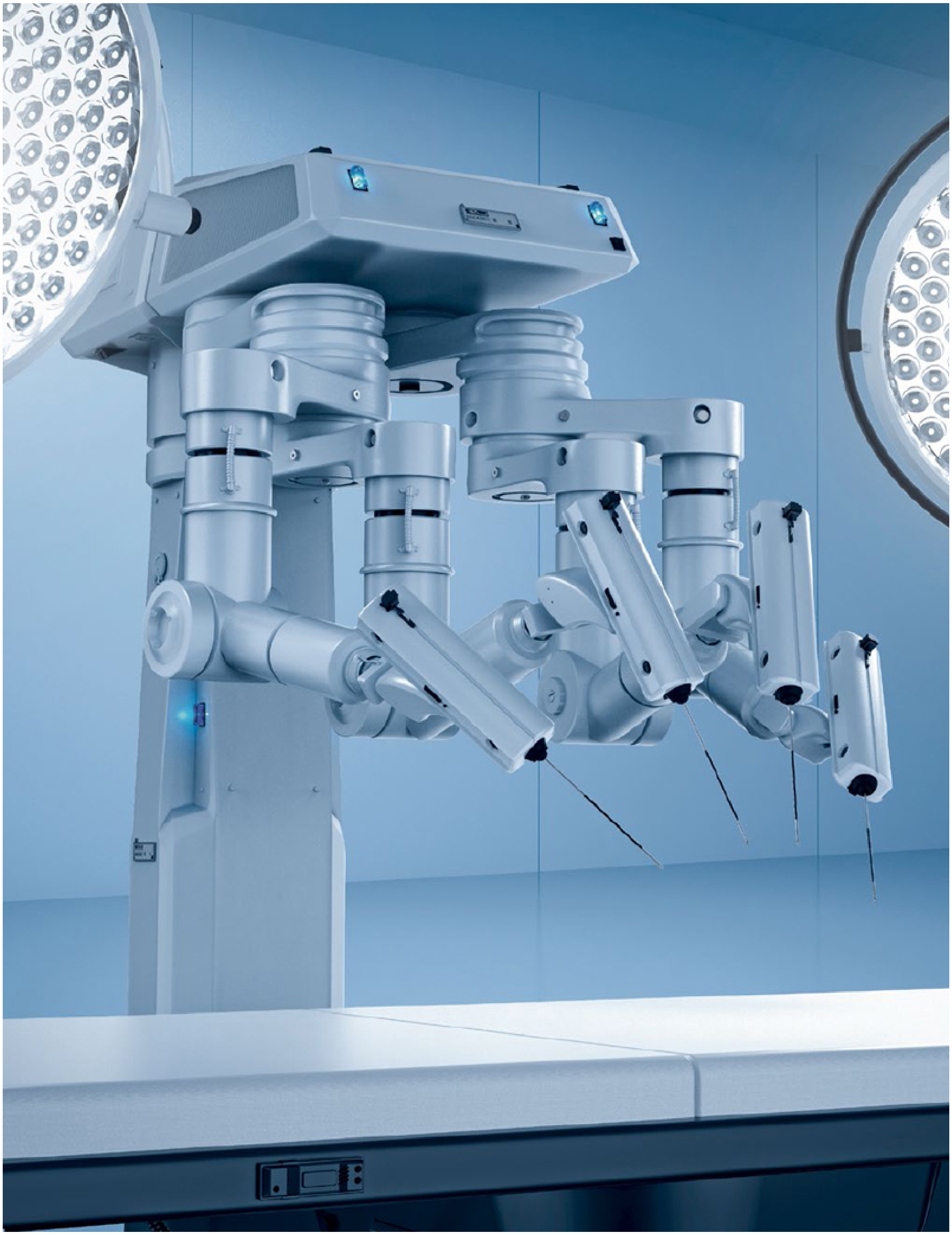
In 2016, the AlphaGo computer program became the first to defeat a professional Go player of the highest rank, without handicaps, when it defeated Lee Sedol (b. 1983) of South Korea. AlphaGo was developed by DeepMind Technologies, a British AI company acquired by Google in 2014. Technically speaking, AlphaGo uses a Monte Carlo tree-search algorithm and artificial neural networks to learn and play the game. In 2017, a newer program version named

AlphaGo Zero learned Go by playing against itself many millions of times, without relying on data from human matches, and then promptly defeated AlphaGo. In some sense, AlphaGo Zero discovered or derived thousands of years of human insight, creativity, and training—and then invented superior approaches in just a few days.

Referring to the amazing playing ap-proaches of AlphaGo, journalist Dawn Chan explains that "from all accounts, one gets the sense that an alien civilization has dropped a cryptic guidebook in our midst: a manual that's brilliant—or at least, the parts of it we can understand."

SEE ALSO Artificial Neural Networks (1943), Mastermind (1970), Backgammon Champion Defeated (1979), Deep Blue Defeats Chess Champion (1997), Othello (1997)

Imagine the future of surgery where autonomous robots play an ever-increasing role, using their own vision systems and machine intelligence. Perhaps they will be the heroes of the operating room as they efficiently draw information from a CT or MRI scan.



2016

AUTONOMOUS ROBOTIC SURGERY

In 2016, the Smart Tissue Autonomous Robot (STAR), a robotic surgical system, demonstrated its skill on a pig's small intestine, using its own enhanced vision, machine intelligence, and dexterity. Compared with human surgeons, STAR's stitches were more consistent and created an intestine that was more resistant to leaks near the seams. In this example of "supervised anatomy," STAR's vision system relied on near-infrared fluorescent tags placed on the intestinal tissue to help its camera track the tissue. STAR planned the suturing task and made adjustments as tissues moved.

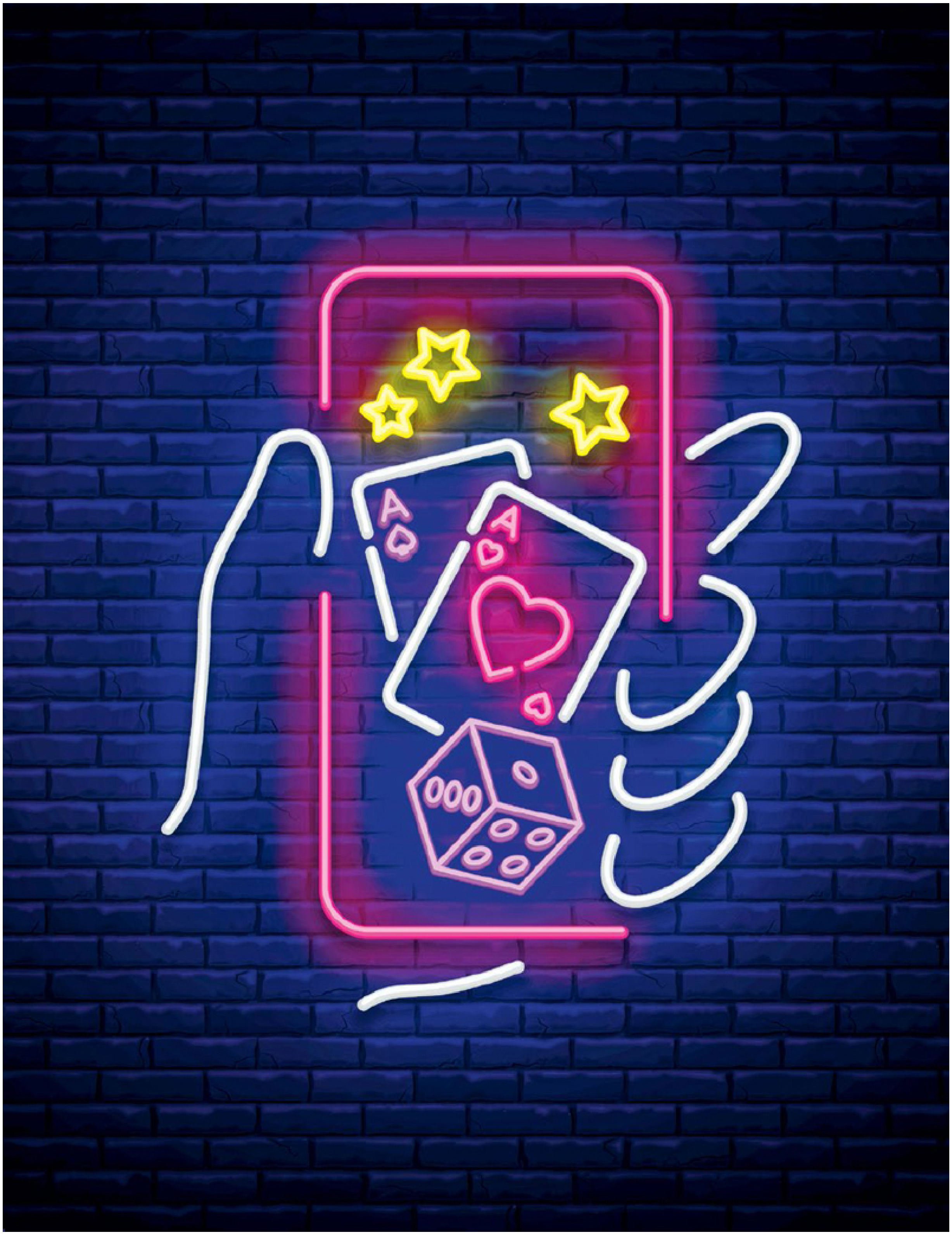
The growing autonomy of surgical robots has had an impressive past, rooted in the increasing use of robots to assist surgery. One of the most popular forms of robotic surgery resembles laparoscopic surgery, also referred to as *keyhole surgery* or *minimally invasive surgery*, which is performed through small incisions and has a reputation for minimizing blood loss and pain while accelerating patient recovery time. However, instead of the surgeon hovering above the patient and directly manipulating the tube-like devices inserted into a patient's body, *robotic surgery* allows the surgeon to sit comfortably at a console and manipulate instruments attached to several robotic arms, while viewing 3-D images transmitted from within the patient's body. Unlike laparoscopic surgery, robotic surgery can suppress the surgeon's hand tremors, and large hand movement can be scaled to provide more accuracy over tiny

movements and manipulations. In the emerging field of telesurgery, robotic instruments connected to high-speed communication networks allow a surgeon to perform an operation on a patient in a separate room.

In 2000, American surgeon Mani Menon (b. 1948) was the first surgeon in the US to use a robot to remove a cancerous prostate gland, and in the same year he established the nation's first center for robotic prostatectomy. Today, robot-assisted laparoscopy is used for hysterectomies, repair of heart mitral valves, hernia repairs, gall bladder removals, and much more. Robots are also used to perform crucial steps in procedures involving orthopedic knee replacements, hair transplants, and Lasik eye surgery.

SEE ALSO Lethal Military Robots (1942), Autonomous Vehicles (1984), AI Death Predictor (2019)

In 2017, AI programs defeated human professional poker players in Texas Hold 'Em. Here, the players have *imperfect information*, which makes the game particularly challenging for computers and requires a kind of "intuition" to determine winning strategies.



2017

AI POKER

In 2017, numerous news stories pro-claimed the huge triumph of two different AI programs that defeated human professional poker players in a game known as Texas Hold ’Em. In the past, AIs have been able to defeat humans in many kinds of games—like chess and Go—involving *perfect information*, where nothing is shielded from the players’ views. In Texas Hold ’Em poker, two or more players are initially dealt two random facedown cards. At the introduction of each new set of public cards, players are asked to “bet, hold, or abandon” the money at stake on the table. Here, the players have *imperfect information*, which makes the game particularly challenging for computers and requires a kind of “intuition” to determine winning strategies. Another challenge is that there is a massive number of possible game scenarios (roughly 10^{160}). In no-limit Hold ’Em, players generally develop betting strategies that play out over numerous hands and often try bluffs (e.g., betting low with favorable cards or betting just to confuse the competition).

In spite of these challenges, an AI named DeepStack was able to defeat professional poker players in heads-up no-limit Texas Hold ’Em. The AI had used deep learning, with millions of randomly generated poker game plays against itself, to train artificial neural networks in order to develop poker intuition ahead of time. Also reported in 2017 was news of another poker AI named Libratus that defeated four of the top-ranked human Texas Hold ’Em players in

numerous games during a twenty-day competition. Libratus did not employ neural networks, instead using a different algorithmic approach called *counterfactual regret minimization* (after every game simulation, the program revisits its decisions and finds ways to improve its strategy). Interestingly, while DeepStack can run on a laptop computer, Libratus required much more sophisticated computer hardware.

Note that AIs that can deal with imperfect information may be useful in numerous real-world situations, like guessing the final sale price of a house or negotiating a good price for a new car. Interestingly, "pokerbots" (poker-playing programs) of various skill levels have been around for years, but they are generally not allowed to be used as assistants to humans in online poker games among humans.

SEE ALSO Artificial Neural Networks (1943), Deep Blue Defeats Chess Champion (1997), Quackle's Scrabble Win (2006), AlphaGo Go Champion (2016)

Researchers have shown that it is possible to place circular psychedelic-looking patches in the field of view of vision AI systems, to fool them into “thinking” a banana is a toaster. This highlights the possible risks of using AI for certain applications.

2018

ADVERSARIAL PATCHES

Imagine a button you could pin to your shirt, or a sticker you could place on a stop sign to trick an AI entity (e.g., a smart surveillance camera or a self-driving car) into thinking you or the sign was anything you wished it to see. Such a scenario is not far-fetched, and it represents a risk for AIs that rely on machine learning, as well as visual or audio systems, to make decisions.

In 2017, researchers at Google designed circular patches with colorful psychedelic patterns to distract AI image classifiers. Such patches can fool an AI system into thinking a banana—or virtually any object—is a toaster, for example, when the patch is placed near the object. Past experiments using other methods have tricked AI systems into thinking that turtles were rifles and that rifles were helicopters. Even though visual adversarial patches can be clearly seen by people, the use of strange patterns (e.g., a graphic on the side of a building or the presence of a complex 3-D sculpture) might be mistaken simply as art, so observers would not realize that the pattern is confusing a drone into thinking a hospital is a military target, for example.

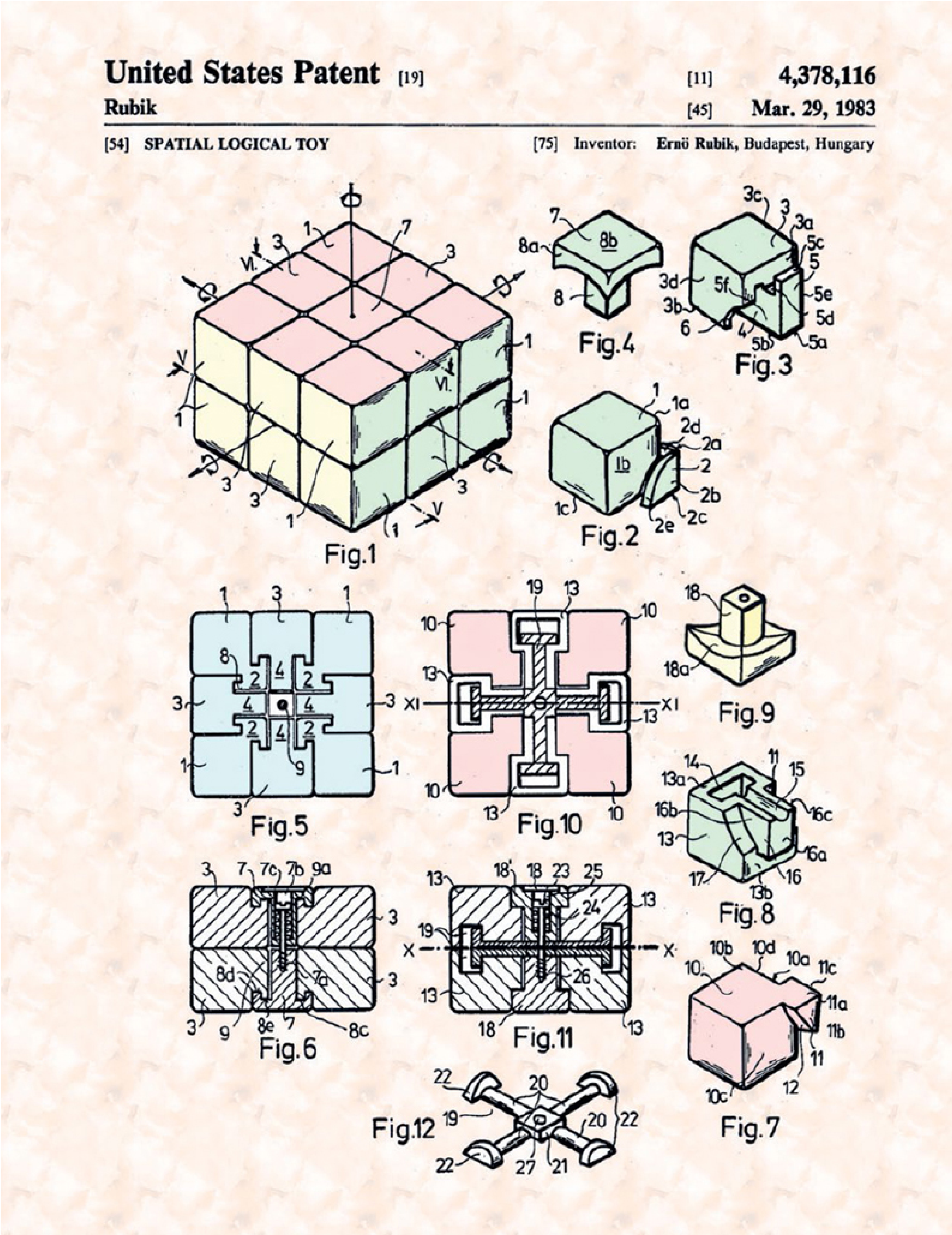
Experiments have also caused AI systems to misclassify stop signs as speed-limit signs. Some past work has focused on changes that are imperceptible to humans, such as changing a few pixels in an image. In 2018, researchers at the University of California, Berkeley, constructed audio adversarial examples for speech

recognition systems. In other words, given any audio waveform, the researchers could produce a nearly identical waveform, tricking speech-to-text systems into transcribing the waveform as any phrase the researchers choose.

"Adversarial machine learning" studies involve manipulating training data while an AI is learning. Although it may be possible to deter some adversaries by requiring AI systems to use multiple classifier systems or to try to program them not to be distracted by adversarial samples during training, potential risks exist in many AI applications.

SEE ALSO Lethal Military Robots (1942), Machine Learning (1959), Ethics of AI (1976), Autonomous Vehicles (1984)

Ernő Rubik’s 1983 US patent 4,378,116 for a “spatial logical toy,” showing internal mechanisms.



2018

RUBIK'S CUBE ROBOTS

The construction of robots capable of solving a Rubik's Cube® using computer vision and physical manipulation has been a popular challenge for AI engineers, who have developed numerous different robotic designs over the years. The original Rubik's Cube was invented by the Hungarian inventor Ernő Rubik (b. 1944) in 1974; by 1982, 10 million of them had been sold in Hungary (which is, curiously, more than the population of that country). It is estimated that over 100 million have been sold worldwide to date.

The cube is a 3 × 3 × 3 array of smaller cubes that are colored in such a way that the six faces of the large cube have six distinct colors. The twenty-six external subcubes are internally hinged so that these six faces can be rotated. The goal of the puzzle is to return a scrambled cube to a state in which each side has a single color. There are 43,252,003,274,489,856,000 different arrangements of the small cubes, and only one of these arrangements is the initial position where all colors match on each of the six sides. If you had a cube for every one of these "legal" positions, you could cover the entire surface of Earth (including oceans) about 250 times.

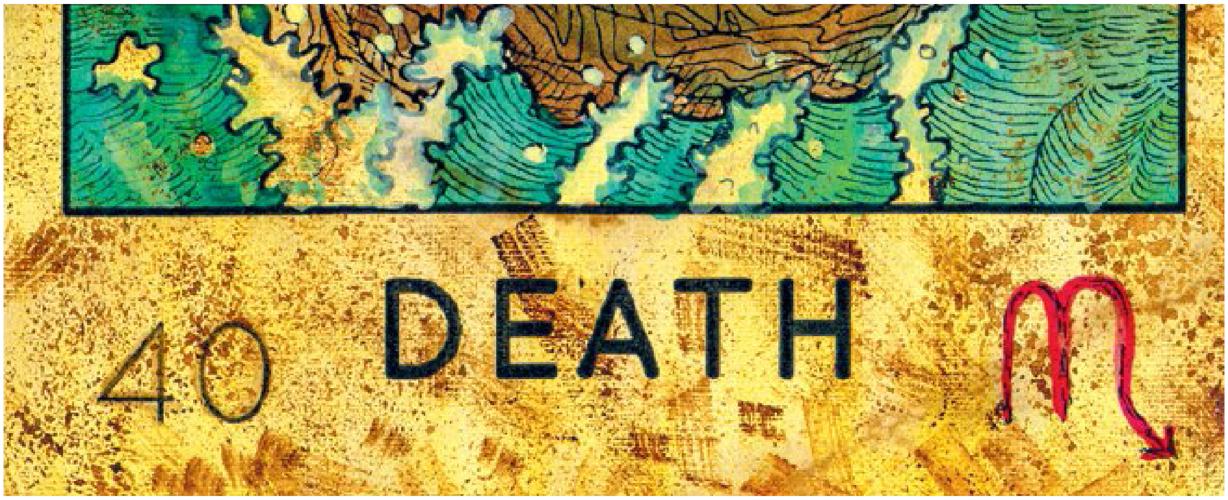
In 2010, researchers proved that no starting configuration of the cube requires more than twenty moves to solve. In 2018, an agile Rubik's Cube robot called the Rubik's Contraption finally pierced the half-second boundary, solving a scrambled cube in just 0.38 seconds, which included image capture, computation, and

movement time. The setup employed by MIT robotics student Ben Katz and software developer Jared Di Carlo made use of six Kollmorgen ServoDisc motors and the so-called Kociemba two-phase algorithm. By comparison, 10.69 seconds was the robotic world record in 2011. Also in 2018, a deep-learning machine finally taught itself how to solve a Rubik's Cube using reinforcement learning without human knowledge.

One variation that never appeared on toy- store shelves is the four-dimensional version of the Rubik's Cube: the Rubik's tesseract. The total number of positions of the Rubik's tesseract is an astounding 1.76×10^{120}. If the rotating layers of either the cube or the tesseract changed positions every second since the beginning of the universe, they would still be turning today and not have exhibited every possible configuration.

SEE ALSO Tower of Hanoi (1883), Reinforcement Learning (1951), Shakey the Robot (1966), ASIMO and Friends (2000)

Researchers have trained an AI system to accurately predict if a person would die in three to twelve months. If you could know the day, or even year, of your death, would you choose to know ahead of time?


40
DEATH

2019

AI DEATH PREDICTOR

In 2016, researchers at Stanford University were able to train an AI system to accurately predict if a person would die in three to twelve months. This remarkable application is included in this book as representative of the wide variety of roles that AI and deep learning will play in coming years.

Palliative care usually involves providing relief for a patient's pain, stress, and other symptoms when the patient has a terminal diagnosis and no cure is expected. Knowing when such specific care is warranted may have beneficial effects for the sufferer, family, and caregivers—and help determine when such care would be most effective. To create the AI "Death Algorithm," the Stanford team used information from about 170,000 patients who had died with, for example, cancer and heart and neurological diseases. Various information from medical records—including a patient's diagnosis, medical procedure, medical-scan codes, drugs prescribed, etc.—was used as input to "teach" the AI system. Then a deep neural net was trained, with various internal weights adjusted for the neuron units. The deep neural net made use of an input layer of 13,654 dimensions (e.g., codes for diagnoses and drugs), 18 hidden layers (each 512 dimensions), and a scalar output layer.

In the end, nine out of ten people predicted to die within three to twelve months *did* die within this time frame. Also, 95 percent of those whom the algorithm determined would outlast twelve months

did live longer lives. However, as physician Siddhartha Mukherjee explained in a recent *New York Times* article: “[The deep learning system] learns, but it cannot tell us why it has learned; it assigns probabilities, but it cannot easily express the reasoning behind the assignment. Like a child who learns to ride a bicycle by trial and error and, asked to articulate the rules that enable bicycle riding, simply shrugs her shoulders and sails away, the algorithm looks vacantly at us when we ask, ‘Why?’ It is, like death, another black box.” Nevertheless, research on these AI death predictors continues, and in 2019 a team of experts at University of Nottingham showed that machine learning could outperform traditional methods at predicting premature deaths, based on demographic, biometric, clinical, and lifestyle factors.

SEE ALSO Deep Learning (1965), Ethics of AI (1976), Autonomous Robotic Surgery (2016)

NOTES AND REFERENCES

"Artificial intelligence is the next step in evolution, but it's a different step. . . . One artificially intelligent device can tell another not only everything it knows in the sense that a human teacher can tell a student some of what he knows, but it can tell another device everything about its own design. . . . Basically the human mind is not most like a god or most like a computer. It's most like the mind of a chimpanzee, and [designed] for getting along in the jungle or out in the fields."

—Edward Fredkin, quoted in Pamela McCorduck's *Machines Who Think*

I've compiled the following reference list that identifies some of the material I used to research and write this book, and to provide information on the quoted sources. As many readers are aware, Internet websites come and go. Sometimes they change addresses or disappear. The website addresses listed here provided valuable background information when this book was written. Online sources such as Wikipedia (en.wikipedia.org) are a valuable starting point for any quest such as this one, and I sometimes used this site as a launchpad, along with many other Websites, books, and research papers.

If I have overlooked an interesting or pivotal moment associated with AI that you feel has not been fully appreciated, please let me know about it. Just visit my website (pickover.com), and send me an email explaining the idea and how you feel it influenced the world. Perhaps future versions of this book will include full-page entries on

AI marvels such as: *Roko’s Basilisk, generative adversarial networks*, *neuromorphic computing*, *Bayesian networks*, *Westworld* (the TV series), *WarGames* (the 1983 film), LSTM (long short-term memory) networks, and more. Finally, I thank my book editors Meredith Hale and John Meils—along with Dennis Gordon, Tom Erickson, Michael Perrone, Teja Krasek, and Paul Moskowitz—for their comments and suggestions.

General Reading

Crevier, D., *AI* (New York: Basic Books, 1993).
Dormehl, L., *Thinking Machines* (New York: Tarcher, 2017).
McCorduck, P., *Machines Who Think* (Natick, MA: A. K. Peters, 2004).
Nilsson, N., *The Quest for Artificial Intelligence* (New York: Cambridge University Press, 2010).
Riskin, J., *The Restless Clock* (Chicago: University of Chicago Press, 2016).
Truitt, E., *Medieval Robots* (Philadelphia: University of Pennsylvania Press, 2015).
Walsh, T., *Machines That Think* (London: C. Hurst & Co., 2017).

Sources

Introduction

Hambling, D., "Lethal logic," *New Scientist*, vol. 236, no. 3151, p. 22, Nov. 11–17, 2017.
Reese, M., "Aliens, Very Strange Universes and Brexit—Martin Rees Q&A," *The Conversation*, April 3, 2017, http://tinyurl.com/mg3w6ez
Truitt, E., *Medieval Robots* (Philadelphia: University of Pennsylvania Press, 2015).
"Visual Trick Has AI Mistake Turtle for Gun," *New Scientist*, vol. 236, no. 3151, p. 19, November 11–17, 2017.

c. 400 BCE: Talos

Haughton, B., *Hidden History: Lost Civilizations, Secret Knowledge, and Ancient Mysteries* (Franklin Lakes, NJ: New Page Books, 2007).

c. 250 BCE: Ktesibios's Water Clock

Dormehl, L., *Thinking Machines* (New York: Tarcher, 2017).

c. 190 BCE: Abacus

Ewalt, D., "No. 2 The Abacus," *Forbes*, August 30, 2005, http://tinyurl.com/yabaocr5
Krimmel, J., "Artificial Intelligence Started with the Calendar and Abacus," 2017, https://tinyurl.com/y5tnoxbl

c. 125 BCE: Antikythera Mechanism

Garnham, A., *Artificial Intelligence: An Introduction* (London: Routledge, 1988).

Marchant, J., “The Antikythera Mechanism: Quest to Decode the Secret of the 2,000-Year-Old Computer,” March 11, 2009, http://tinyurl.com/ca8ory

1206: Al-Jazari’s Automata

Hill, D., *Studies in Medieval Islamic Technology*, ed. D. A. King (Aldershot, Great Britain: Ashgate, 1998).

c. 1220: Lancelot’s Copper Knights

Riskin, J., *The Restless Clock* (Chicago: University of Chicago Press, 2016).
Truitt, E., *Medieval Robots* (Philadelphia: University of Pennsylvania Press, 2015).

c. 1300: Hesdin Mechanical Park

Bedini, S., “The Role of Automata in the History of Technology,” *Technology and Culture*, vol. 5, no. 1, pp. 24–42, 1964.
Lightsey, S., *Manmade Marvels in Medieval Cultures and Literature* (New York: Palgrave, 2007).

c. 1305: Ramon Llull’s *Ars Magna*

Dalakov, G., “Ramon Llull,” http://tinyurl.com/ybp8rz28
Gray, J., “‘Let us Calculate!’: Leibniz, Llull, and the Computational Imagination,” http://tinyurl.com/h2xjn7j
Gardner, M., *Logic Machines and Diagrams* (New York: McGraw-Hill, 1958).
Madej, K., *Interactivity, Collaboration, and Authoring in Social Media* (New York: Springer, 2016).
Nilsson, N., *The Quest for Artificial Intelligence* (New York: Cambridge University Press, 2010).

1352: Religious Automata

Coe, F., *The World and Its People, Book V, Modern Europe*, ed. L. Dunton (New York: Silver, Burdett, 1896).
Fraser, J., *Time, the Familiar Stranger* (Amherst: University of Massachusetts Press, 1987).

c. 1495: Leonardo da Vinci’s Robot Knight

Phillips, C., and S. Priwer, *The Everything Da Vinci Book* (Avon, MA: Adams Media, 2006).
Rosheim, M., *Leonardo’s Lost Robots* (New York: Springer, 2006).

1580: Golem

Blech, B., "Stephen Hawking's Worst Nightmare? Golem 2.0" (tagline), *The Forward*, January 4, 2015, http://tinyurl.com/yats534k

1651: Hobbes's *Leviathan*

Dyson, G., *Darwin among the Machines* (New York: Basic Books, 1997).

1714: The Consciousness Mill

Bostrom, N., "The Simulation Argument: Why the Probability that You Are Living in a Matrix is Quite High." *Times Higher Education Supplement*, May 16, 2003, http://tinyurl.com/y8qorjcf

Moravec, H., "Robot Children of the Mind." In David Jay Brown's *Conversations on the Edge of the Apocalypse* (New York: Palgrave, 2005).

1726: Lagado Book-Writing Engine

Weiss, E., "Jonathan Swift's Computing Invention." *Annals of the History of Computing*, vol. 7, no. 2, pp. 164–165, 1985.

1738: de Vaucanson's Duck Automaton

Glimcher, P., *Decisions, Uncertainty, and the Brain: The Science of Neuroeconomics*. (Cambridge, MA: MIT Press, 2003).

Riskin, J., "The Defecating Duck, or, the Ambiguous Origins of Artificial Life." *Critical Inquiry*, vol. 29, no. 4, pp. 599–633, 2003.

1770: Mechanical Turk

Morton, E., "Object of Intrigue: The Turk, a Mechanical Chess Player that Unsettled the World." August 18, 2015, http://tinyurl.com/y72aqfep

1774: Jaquet-Droz Automata

Lorrain, J., *Monsieur De Phocas* (trans. F. Amery) (Sawtry, Cambridgeshire, UK: Dedalus, 1994).

Riskin, J., *The Restless Clock* (Chicago: University of Chicago Press, 2016).

1818: *Frankenstein*

D'Addario, D., "The Artificial Intelligence Gap Is Getting Narrower," *Time*, October 10, 2017, http://tinyurl.com/y8g5bu5o

Gallo, P., " Are We Creating a New Frankenstein?" *Forbes*, March 17, 2017, http://tinyurl.com/ycsdr6gt

1821: Computational Creativity

Colton, S., and G. Wiggins, "Computational Creativity: The Final Frontier?" In *Proceedings of the 20th European Conference on Artificial Intelligence*, 2012.

1854: Boolean Algebra

Titcomb, J., "Who is George Boole and Why is He Important?" *The Telegraph*, November 2, 2015, http://tinyurl.com/yb25t8ft

1863: "Darwin among the Machines"

Wiener, N., "The Machine Age," 1949 unpublished essay for the *New York Times*, http://tinyurl.com/ybbpeydo

1868: *The Steam Man of the Prairies*

Liptak, A., "Edward Ellis and the Steam Man of the Prairie," *Kirkus*, November 6, 2015, http://tinyurl.com/yadhxn7t

1907: Tik-Tok

Abrahm, P., and S. Kenter, "Tik-Tok and the Three Laws of Robotics," *Science Fiction Studies*, vol. 5, pt. 1, March 1978, http://tinyurl.com/ybm6qv2y

Goody, A., *Technology, Literature and Culture* (Malden, MA: Polity Press, 2011).

1920: *Rossum's Universal Robots*

Floridi, L., *Philosophy and Computing* (New York: Taylor & Francis, 2002).

Stefoff, R., *Robots* (Tarrytown, NY: Marshall Cavendish Benchmark, 2008).

1927: *Metropolis*

Lombardo, T., *Contemporary Futurist Thought* (Bloomington, IN: AuthorHouse, 2008.)

1942: Asimov's Three Laws of Robotics

Markoff, J., "Technology: A Celebration of Isaac Asimov," *New York Times*, April 12, 1992, http://tinyurl.com/y9gevq6t

1943: Artificial Neural Networks

Lewis-Kraus, G., "The Great A.I. Awakening," *New York Times Magazine*, December 14, 2016, http://tinyurl.com/gue4pdh

1950: *The Human Use of Human Beings*

Crevier, D., *AI* (New York: Basic Books, 1993).
Wiener, N., *The Human Use of Human Beings* (London: Eyre & Spottiswoode, 1950).

1952: Speech Recognition

“Now We’re Talking: How Voice Technology is Transforming Computing,” *The Economist*, January 7, 2017, http://tinyurl.com/yaedcvfg

1954: Natural Language Processing

“701 Translator,” IBM Press Release, January 8, 1954, http://tinyurl.com/y7lwblng

1956: Dartmouth AI Workshop

Dormehl, L., *Thinking Machines* (New York: Tarcher, 2017).
McCorduck, P., *Machines Who Think* (Natick, MA: A. K. Peters, 2004).

1957: Transhumanism

Huxley, J., *New Bottles for New Wine* (London: Chatto & Windus, 1957).
Istvan, Z., “The Morality of Artificial Intelligence and the Three Laws of Transhumanism,” *Huffington Post*, http://tinyurl.com/ycpx9bwa
Pickover, C., *A Beginner’s Guide to Immortality* (New York: Thunder’s Mouth Press, 2007).

1959: Knowledge Representation and Reasoning

Nilsson, N., *The Quest for Artificial Intelligence* (New York: Cambridge University Press, 2010).

1964: ELIZA Psychotherapist

Weizenbaum, J., “ELIZA—A Computer Program for the Study of Natural Language Communication Between Man and Machine,” *Communications of the ACM*, vol. 9, no. 1, pp. 36–45, 1966.

1964: Face Recognition

West, J., “A Brief History of Face Recognition,” http://tinyurl.com/y8wdqsbd

1965: Expert Systems

Dormehl, L., *Thinking Machines* (New York: Tarcher, 2017).

1965: Fuzzy Logic

Carter, J., "What is 'Fuzzy Logic'?" *Scientific American*, http://tinyurl.com/yd24gngp

1965: Deep Learning

Fain, J., "A Primer on Deep Learning," *Forbes*, December 18, 2017, http://tinyurl.com/ybwt9qp3

1967: Living in a Simulation

Davies, P., "A Brief History of the Multiverse," *New York Times*, 2003, http://tinyurl.com/y8fodeoy.

Koebler, J., "Is the Universe a Giant Computer Simulation?" http://tinyurl.com/y9lluy7a

Reese, M., "In the Matrix," http://tinyurl.com/y9h6fjyx.

1972: Paranoid PARRY

Wilks, Y., and R. Catizone, "Human-Computer Conversation," arXiv:cs/9906027, June 1999, http://tinyurl.com/y7erxtxm

1975: Genetic Algorithms

Copeland, J., *The Essential Turing* (New York: Oxford University Press, 2004).

Dormehl, L., *Thinking Machines* (New York: Tarcher, 2017).

1979: Backgammon Champion Defeated

Crevier, D., *AI* (New York: Basic Books, 1993).

1982: *Blade Runner*

Guga, J., "Cyborg Tales: The Reinvention of the Human in the Information Age," in *Beyond Artificial Intelligence* (New York: Springer, 2015).

Littman, G., "What's Wrong with Building Replicants?" in *The Culture and Philosophy of Ridley Scott* (Lanham, MD: Lexington).

1984: Autonomous Vehicles

Lipson, H., and M. Kurman, *Driverless* (Cambridge, MA: MIT Press, 2016).

1986: Swarm Intelligence

Jonas, David, and Doris Jonas, *Other Senses, Other Worlds* (New York: Stein and Day, 1976).

1988: Moravec's Paradox

Elliott, L., "Robots Will Not Lead to Fewer Jobs—But the Hollowing Out of the Middle Class." *The Guardian*, August 20, 2017, http://tinyurl.com/y7dnhtpt

Moravec, H., *Mind Children* (Cambridge, MA: Harvard University Press, 1988).

Pinker, S., *The Language Instinct* (New York: William Morrow, 1994).

1990: "Elephants Don't Play Chess"

Brooks, R., "Elephants Don't Play Chess," *Robotics and Autonomous Systems,* vol. 6, pp. 139–159, 1990.

Shasha, D., and C. Lazere, *Natural Computing* (New York: Norton, 2010).

1993: Leakproof "AI Box"

Readers may wish to become familiar with the concept of *Roko's Basilisk*, a thought experiment in which future AI systems retaliate against people who did not bring the AI systems into existence. In many versions of Roko's Basilisk, AIs retroactively punish people by torturing simulations of these people.

Vinge, V., "The Coming Technological Singularity." VISION-21 Symposium, March 30–31, 1993.

1994: Checkers and AI

Madrigal, A., "How Checkers Was Solved." *The Atlantic*, July 19, 2017, http://tinyurl.com/y9pf9nyd

1997: Deep Blue Defeats Chess Champion

Webermay, B., "Swift and Slashing, Computer Topples Kasparov." May 12, 1997, http://tinyurl.com/yckh6xko

2000: ASIMO and Friends

Wiener, N., *God and Golem* (Cambridge, MA: MIT Press, 1964).

2001: Spielberg's *A.I. Artificial Intelligence*

Gordon, A., *Empire of Dreams: The Science Fiction and Fantasy Films of Steven Spielberg* (New York: Rowman & Littlefield, 2007).

2002: Roomba

Reel, M., "How the Roomba Was Realized." *Bloomberg*, October 6, 2003, http://tinyurl.com/yd4epat4

2002: Solving the Game of Awari

Romein, J., and H. Bal, “Awari is Solved.” *ICGA Journal*, September 2002, pp. 162–165.

2003: Paperclip Maximizer Catastrophe

It should be noted that AI researcher Eliezer Yudkowsky (b. 1979) has said that the paperclip maximizer idea may have originated with him. See the podcast “Waking Up with Sam Harris #116—AI: Racing Toward the Brink” (with Eliezer Yudkowsky).

2006: Quackle’s Scrabble Win

Anderson, M., “Data Mining Scrabble.” *IEEE Spectrum*, vol. 49, no. 1, p. 80.

2011: Watson on *Jeopardy!*

Jennings, K., “My Puny Human Brain.” *Slate*, Feb. 16, 2011, http://tinyurl.com/86xbqfq

2015: “Call Them Artificial Aliens”

Kelly, K., “Call them Artificial Aliens,” in Brockman, J., ed., *What to Think About Machines That Think* (New York: Harper, 2015).

2015: AI on Mars

Fecht, S., “The Curiosity Rover and Other Spacecraft Are Learning to Think for Themselves.” *Popular Science*, June 21, 2017, http://tinyurl.com/y895pq6k,

Koren, M., “The Mars Robot Making Decisions On Its Own,” *The Atlantic*, June 23, 2017, http://tinyurl.com/y8s8alz6

2016: AlphaGo Go Champion

Chan, D., “The AI That has Nothing to Learn from Humans.” *The Atlantic*, October 20, 2017. http://tinyurl.com/y7ucmuzo

Ito, J., and J. How, *Whiplash: How to Survive Our Faster Future* (New York: Grand Central Publishing, 2016).

2018: Adversarial Patches

Brown, T., et al., “Adversarial Patch,” 31st Conference on Neural Information Processing Systems (NIPS), Long Beach, CA, 2017.

2019: AI Death Predictor

Avati, A., et al., “Improving Palliative Care with Deep Learning,” *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Kansas City, MO, pp. 311–316, 2017.

Mukherjee, S., “This Cat Sensed Death. What if Computers Could, Too?” *New York Times*, January 3, 2018, http://tinyurl.com/yajko6pv

Rajkomar, A., et al., “Scalable and accurate deep learning with electronic health records,” *npj Digital Medicine,* vol. 1, no. 18, 2018, http://tinyurl.com/ych74oe5

IMAGE CREDITS

Because several of the old and rare illustrations shown in this book were difficult to acquire in a clean and legible form, I have sometimes taken the liberty to apply image-processing techniques to remove dirt and scratches, enhance faded portions, and occasionally add a slight coloration to a black-and-white figure in order to highlight details or to make an image more compelling. I hope historical purists will forgive these slight artistic touches and understand that my goal was to create an attractive book that is aesthetically interesting for a wide audience. My fascination for the incredible depth and diversity of topics revolving around AI should be evident through the photographs and drawings.

Alamy: The Advertising Archives: 72; David Fettes: 154; richterfoto: 152; Science Photo Library: 128

Courtesy of International Business Machines Corporation, © (1962) International Business Machines Corporation: 88

Courtesy of Universal Studios Licensing LLC: endpapers, 126

Cyclopaedia: Abraham Rees: 6

Gallica: Robert de Boron: 14; Guillaume de Machaut: 16

Internet Archive: J.J. Grandville: 30; Ramon Llull: 18

iStock/Getty Images Plus: ChubarovY: 195; in-future: 195; Kickimages: 138; Paulbr: 158; Vera Petruk: 198; PhonlamaiPhoto: 190; JIRAROJ PRADITCHAROENKUL: 76; Rouzes: 110; sergeyryzhov: 8; Soifer: 192; Vladimir Timofeev: 156

Getty: Al Fenn: 90; Kyodo News/Contributor: 168; The Washington Post/Contributor: 180; Westend61: 82

Google Research Team: 194

Library of Congress: Charles Verschuuren: 64

New York World's Fair 1939-1940 records, Manuscripts and Archives Division, The New York Public Library/Mansfield Memorial Museum: endpapers, 68

Scientific American: 136

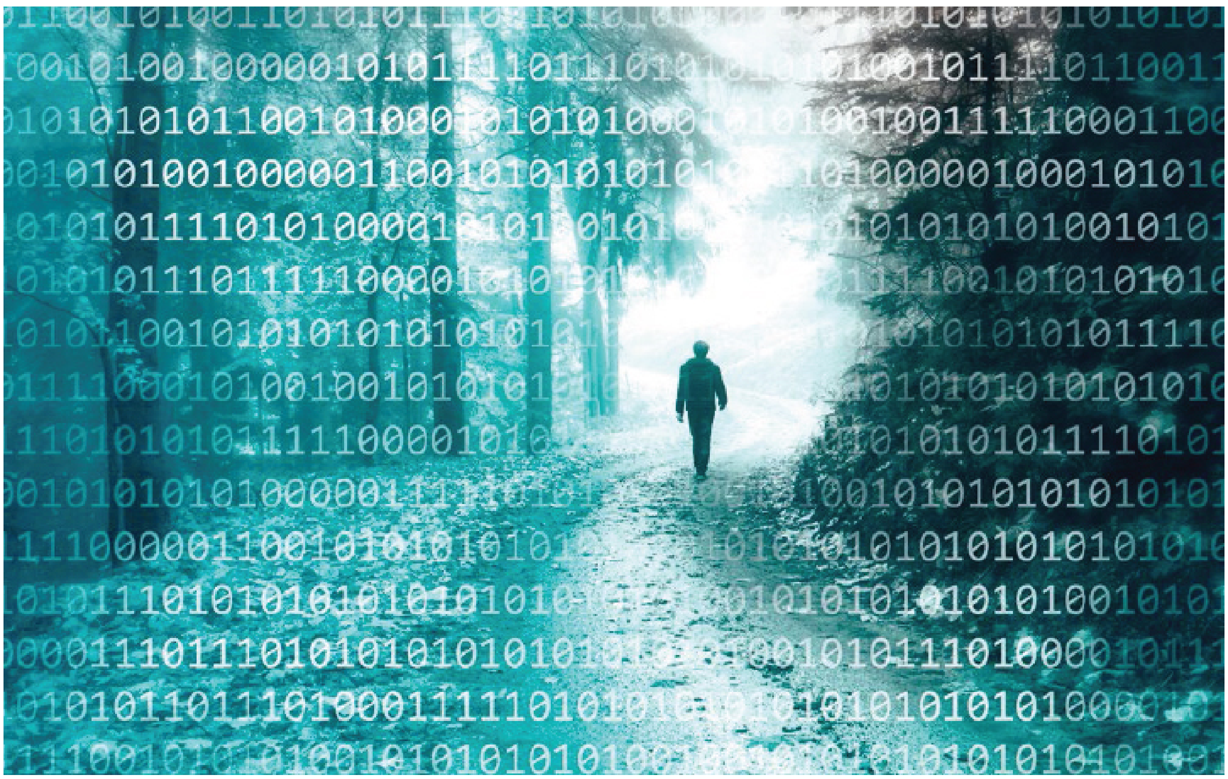
Shutterstock: 7th Son Studio: 146; Charles Adams: 124; Berke: 170; Black Moon: 44; gualtiero boffi: 122; Willyam Bradberry: endpapers, 48; camilla$$: 178; Chesky: 142; Esteban De Armas: 74; Digital Storm: 184; Dmitry Elagin: 52; Leonid Eremeychuk: 134; Evannovostro: 114; Everett Historical: 78; Ilterriorm: 166; Eugene Ivanov: back cover, 24; Ala Khviasechka: xii; Anastasiia Kucherenko: 120; MicroOne: 86; Morphart Creation: 26; MossStudio: 98; NadyaEugene: 150; Nor Gal: 164; Ociacia: back cover, 96; Phonlamai Photo: 60, 84, 108; Photobank gallery: 46; PHOTOCREO/Michal Bednarek: endpapers, 102; Saran_Poroong: 188; Glenn Price: 176; Quality Stock Arts: 174; R-Type: 130; ra2studio: 140; Robsonphoto: 212; Benjavisa Ruangvaree: 28; Tatiana Shepeleva: 100; Travel mania: 148; Trial: 172; Tsuneomp: 144; videodoctor: 118

Stanford University: Chuck Painter: 92

United States Patent and Trademark Office: 56, 62, 106, 112, 160, 196

Wikimedia: Rocky Acosta: 80; Ismail al-Jazari/MFA Boston: 12; Aquarius70: 162; Edward S. Ellis/Cover of “Steam Man of the Prairies”: 50; Emesee/Electric Sheep: 40; David Iliff: 20; Jastrow/Ludovisi Collection: 4; Erik Möller, Mensch-Erfinder-Genie exhibit, Berlin: 22; Jessica Mullen: 182; NASA: back cover, 94, 132, 136; John R. Neill: 58; New York World’s Fair 1939: 70; Joseph Racknitz/Humboldt University Library: 34; Rama: 36; William Bruce Ellis Ranken/Christie’s: 104; SRI International: 116; Sybil Tawse/“Bulfinch’s Mythology”: 2; Robert Toombs/New York Five Cent Library: 54; Carston Ullrich: back cover, 42; Mogi Vicentini: 10; Theodore Von Holst: 38; Jiuguang Wang: 66

ABOUT THE AUTHOR



Clifford A. Pickover is a prolific author, having published more than fifty books, translated into over a dozen languages, on topics ranging from science and mathematics to religion, art, and history. He received his PhD degree from Yale University, has been granted more than 600 U.S. patents, and has more than 34,000 Twitter followers. His patents are owned by a wide variety of companies, including IBM, Ebay, Google, Twitter, Yahoo!, Paypal, LinkedIn, and many more. His website, Pickover.com, has had millions of visitors.

When describing his work, creativity, and sense of wonder, *The New York Times* wrote, “Pickover contemplates realms beyond our known reality.” According to *WIRED*, “Bucky Fuller thought big, Arthur C. Clarke thinks big, but Cliff Pickover outdoes them both.” The *Christian Science Monitor* proclaimed: “Pickover inspires a new generation of da Vincis to build unknown flying machines and create new Mona Lisas.”